APA Newsletter on

Philosophy and Computers

Piotr Bołtuć, Editor | Fall 2008 | Volume 08, Number 1

From the Editor

Piotr Bołtuć
University of Illinois at Springfield

My goal two years ago, when I became the editor of this *Newsletter*, was to demonstrate the relevance of the topics pertinent to the issues of *philosophy and computers* for main stream, top level philosophy. This was accomplished with publication of articles, and discussion papers, in the areas of metaphysics (e.g., Baker, Wheeler, and Thomasson), epistemology (e.g., Harman, Rapaport, and Franklin/Baars/Ramamurthy), and moral theory (e.g., Moor, Floridi, and Bynum).

My second goal was to highlight the potential, and accomplishments, of online education in philosophy. We have presented some such accomplishments, including the program at British Open University. In the following couple of issues we shall present online philosophy programs at The University of Illinois at Springfield, the American Public University and other institutions.

In the following issues we intend to continue on with those two areas of focus, while giving also more attention to applied topics. The philosophical relevance of *computers* is not limited to the niche-area of some sort of applied philosophy, though applied philosophy is indeed a part of what is pertinent to our mission.

The next issue (08:2) will be focused in part on autonomy and responsibility of robotic agents, with papers by D. Berkich and J. Sullins among others. The following issue (09:1) is a joint project of this *Newsletter* and the Newsletter of the American Society for Aesthetics devoted to the philosophical topics in computer art, with papers by Dom McIver Lopes and Derek Matravers and many others. Now, informed on the plans of this *Newsletter* for the future, and of its past, the Reader deserves introduction to the current edition.

**

The Global Workspace Theory is often viewed as the leading cognitive theory of consciousness; it is a theory of *functionally consciousness*, as the authors put it. It is less clear what the leading scientific theory of phenomenal consciousness is (the *dialogue of hemispheres* hypothesis is a decent candidate). The issue is hard. In their featured article Franklin, Baars, and Ramamurthy claim that adding a mechanism to stabilize the perceptual field "might provide a significant step toward phenomenal consciousness in machines." It seems to me that such a mechanism would enhance quality of phenomenal consciousness, if the latter was already present (this way the step may indeed be significant), but the question whether the mechanism would help explain, or produce, phenomenal consciousness remains open. This point is the flip-side of the issue raised by Gilbert Harman. Harman, in his brief commentary to the Franklin, Baars, and Ramamurthy article, asks how perceptual stability is provided by *phenomenal consciousness*; we also have a reply by Baars et al. The article is bound to produce much further discussion.

Franklin, Baars, and Ramamurthy refer to the new peer-reviewed journal of machine consciousness that is "soon to be published." We are glad now to publish an article by the editor of *The International Journal of Machine Consciousness* (in statu nascendi), Antonio Chella. While Franklin et al. view stabilizing of perceptual field as a step towards machine consciousness, Chella focuses, for the same purpose, on a generalized loop among the brain, the body, and the environment. Such loop comprises the interactions between the robot and the environment viewed as "the perception experience of the robot." The article is a nice follow up on another paper proposing the externalist theory of consciousness, by R. Manzotti, published by this *Newsletter* (06:2); it is also a good fit with a web-related sort of externalism in the work of H. Halpin, also here (see 07:1; commentary by M. Wheeler in the current issue).

The topic of machine consciousness is also relevant for the following two papers: Harman's, where he responds to his commentators (as well as becoming a commentator to the current featured article), and Wheeler's. Michael Wheeler emphasizes the exceptional quality of Harry Halpin's project. He points out that Halpin proposes philosophy of the web as the fourth conceptual anchor of the notion of mind, apart from, and apart with, the classical, connectionist, and embodied-embedded models. Wheeler also presents a set of answers to Halpin's criticism of his work.

The next group of articles pertains to Luciano Floridi's "Understanding Information Ethics" featured by us last year (06:2). Terry Bynum, who does not have to be introduced here as one of the pioneers of computer ethics, places Floridi's work in the broad historical context and compares his *Platonic-Spinozian* approach to a materialist approach of Wiener. John Barker in his neat analytical paper argues that "overall complexity, or quantity of information" can hardly be defined since the notion is essentially context-dependent and the context is always pragmatic. This would be detrimental to Floridi's ethics guided by a non-anthropomorphic directive to preserve information. E.H. Spence puts Floridi's proposal against the backdrop of other moral theories, especially that by Gewirth.

The following group of papers comment on Lynne Baker's article "The Shrinking Difference between Artifacts and Natural Objects" (07:1). Amie Thomasson supports Baker's thesis that artifacts should be considered a genuine part of our world. She argues against the requirement for them to be definable in a mind-independent manner. Beth Preston argues that Baker's rejection of the distinction between artifacts and natural objects, while going in the right direction, does not go quite far enough.

Preston shows that, in ontology, the whole distinction between intention-dependent and intention-independent objects cannot be maintained. Peter Kroes and Pieter Vermaas agree with Baker in part and disagree in part. They claim that artifacts are quite different from natural objects, but not quite inferior to them.

In the last part of the *Newsletter* we are happy to publish a book review of Amie Thomasson's recent book *Ordinary Objects* by Huaping Lu-Adler as well as three papers pertaining to various aspects of computers in education. Harriet Baber argues against ordering published anthologies. It is more cost-efficient for students, and in many cases quite convenient for the instructor, to find all pertinent information online. Vince Müller and Gordana Dodig-Crnkovich in their respective papers discuss various aspects of the international course on *Information Ethics* organized by those two authors as well as Gaetano Lanzarone, Keith Miller, and myself, with further support from Bill Rapaport, Marvin Croy, and Luciano Floridi. The course is offered, for the first time, online, from the campuses in Sweden, USA, Italy, and Greece in the Fall of 2008.

We close with two notes. The first one, by Constantinos Athanasopoulos, presents a conference on e-Learning in philosophy and related disciplines conducted in Scotland this May. The second note is a brief reminder of our special interest in exploring the question of *the ontological status of web-based objects*.

***

The present issue is dominated by discussion articles pertaining to the papers we published in the past, which I am very happy about. It is easy in philosophy to talk past each other but harder, though by far more rewarding, to engage in a true conversation. Once again, I want to thank those who make it possible for me to help animate this conversation as the editor here: Dean Margot Duley at UIS, my departmental colleagues, Bekeela Watson (my editorial intern), and, last but not least, the chair and members of this Committee.

A few months ago David Chalmers came up with the idea of turning this newsletter into a journal, a nice idea *prima facie*. There is some support for it with the Committee, and with the APA, though nobody is rushing us into anything. And this is good; after all, it is better to have a newsletter that is sort of like a journal than a journal that is like a newsletter.

From the Chair

Michael Byron
Kent State University

The Pacific Division Meeting, held March 19-23 in Pasadena, was an interesting one for the Committee. Last fall, the Committee voted to award the Barwise Prize to David Chalmers. The Committee hosted a special session to award the prize, at which Chalmers spoke. His presentation explored the "extended mind thesis," according to which it is possible for cognitive systems to exist outside the head, and argued that his iPhone represented just such an extension.

The Committee also sponsored a special session at the Pacific Meeting. The session, ably chaired by newsletter editor Peter Boltuc, presented "Pedagogical Developments in Philosophy and Computers." Participants included Patrick Suppes of Stanford, "Introducing Gifted Elementary-School Students to Formal Proofs," Peter Boltuc from the University of Illinois–Springfield, on "Teaching Philosophy Online: Beyond Logic," and Marvin Croy of the University of North Carolina–Charlotte, on "Using Educational Data Mining to Provide Hints for Proof Construction."

The Committee looks forward to another productive year in 2008-09.

Papers on Robot Consciousness

Featured Article

A Phenomenally Conscious Robot?

Stan Franklin
University of Memphis

Bernard J. Baars
The Neuroscience Institute, San Diego

Uma Ramamurthy
St. Jude Children's Research Hospital

Abstract

The currently leading cognitive theory of consciousness, Global Workspace Theory (Baars 1988 and 2003), postulates that the primary functions of consciousness include a global broadcast serving to recruit internal resources with which to deal with the current situation and to modulate several types of learning. In addition, conscious experiences present current conditions and problems to a "self" system, an executive interpreter that is identifiable with brain structures like the frontal lobes and precunues (Baars 1988). Be it human, animal, or artificial, an autonomous agent (Franklin and Graesser 1997) is said to be functionally consciousness *if its control structure (mind) implements Global Workspace Theory and the LIDA Cognitive Cycle, which includes unconscious memory and control functions needed to integrate the conscious component of the system. We would therefore consider humans, many animals (Seth, Baars, and Edelman 2005) and even some virtual or robotic agents (Franklin 2003; Shanahan 2006) to be functionally consciousness. Such entities may approach phenomenal consciousness as additional brain-like features are added. Here we argue that adding mechanisms to produce a stable, coherent perceptual field (Merker 2005) in a LIDA controlled mobile robot might provide a significant step toward phenomenal consciousness in machines (Franklin, 2005).*

Machine Consciousness

In the last decade there have been increasing efforts to address the question of machine consciousness. A number of computational models have been proposed and implemented, international conferences have been held, and a peer-reviewed journal will soon be published.

A 2001 workshop entitled "Can a machine be conscious?" (http://www.theswartzfoundation.org/banbury_e.asp) was the impetus for a community of researchers to embark on the serious, scientific study of the possibility of machine consciousness. This was followed by subsequent such workshops in Torino, Italy (2003) (http://jacob.disam.etsii.upm.es/public/events/mcc/), Lesvos, Greece (2005) (http://www.icsc-naiso.org/conferences/bics2006/bics06-cfp.html), and Espoo, Finland (2008) (http://www.stes.fi/step2008/program.

html). At these meetings various projects aimed at eventually achieving machine consciousness were reported on. These include Igor Aleksander's MAGNUS (2000), Rodney Cotterill's CyberChild (2003), Owen Holland's CRONOS (2007), Pentti Haikonen's Cognitive Machine (2007), Stan Franklin's LIDA (Franklin and Patterson 2006), and others.

Functional vs. Phenomenal Consciousness

The currently leading cognitive theory of consciousness, Global Workspace Theory (Baars 1988 and 2003), postulates that the primary functions of consciousness include a global broadcast serving to recruit internal resources with which to deal with the current situation and to modulate several types of learning. In addition, conscious experiences present current situations and problems to a "self" system, an executive interpreter that is identifiable with brain structures like the frontal lobes and precunues (Baars 1988). Be it human, animal, or artificial, an agent (Franklin and Graesser 1997) is said to be *functionally consciousness* if its control structure (mind) implements Global Workspace Theory and the LIDA Cognitive Cycle, which includes unconscious memory and control functions needed to integrate the conscious component of the system. We would consider humans, many animals (Seth, Baars, and Edelman 2005), and even some virtual or robotic agents (Franklin 2003; Shanahan 2006) to be functionally consciousness.

We must carefully distinguish functional consciousness from the usual use of "consciousness," which assumes phenomenal experience, the subjective experience of qualia. To keep this distinction clear we will refer to consciousness in this usual usage as *phenomenal consciousness*. The machine consciousness projects mentioned above are all aimed at eventually achieving artificial phenomenal consciousness. Is this even possible? We believe that an embodied robotic version of LIDA—which would meet a number of criteria for human consciousness—will perhaps be the closest entity to artificial phenomenal consciousness. Phenomenal consciousness is argued to exist in biological entities that have a sizable set of known features (Seth et al. 2005; Baars 1988). As functionally conscious computational entities meet more and more of these criteria, machine conscious robots may become indistinguishable, as to consciousness, from biologically conscious animals. (Please see http://consc.net/mindpapers/6.1d for an exhaustive list of articles on the subject.)

The IDA Software Agent

IDA (Intelligent Distribution Agent) is an intelligent software agent (Franklin and Graesser 1997) developed for the U.S. Navy (Franklin et al. 1998). While IDA was developed for a specific set of human tasks, it reflects broader principles of human cognition. In the initial IDA implementation, its aim was to simulate human "detailers," whose job it is to assign U.S. Navy sailors to suitable jobs. At the end of each sailor's tour of duty, he or she is assigned to a new billet. This complex assignment process is called distribution. The Navy employs some 300 trained people, called detailers, full time to effect these new assignments. IDA facilitates this process by completely automating the role of the human detailer (Franklin 2001). Communicating with sailors by email in unstructured English, IDA negotiates with them about new jobs, employing constraint satisfaction, deliberation and volition, eventually assigning a job constrained by both human and organizational requirements. The IDA software agent is currently up and running and has matched the performance of the Navy's human detailers.

IDA is quite a complex software agent (Franklin and Graesser 1999) that models a broad swath of human cognition including "consciousness" in the sense of implementing Global Workspace Theory. IDA exhibits both external and internal voluntary action selection, as well as consciously mediated action selection of both the internal and external variety. She uses her "consciousness" module to handle routine problems with novel content. It also allows her to watch for unexpected events—both dangers and opportunities. All this together makes a strong case, in our view, for functional consciousness as defined above.

But, is IDA phenomenally consciousness? We have argued earlier that there are "no convincing arguments for such a claim" (Franklin 2003) and currently see no reason to change that view. It seems that IDA implements part, but not all, of consciousness. What needs to be added to an IDA-based software agent to achieve phenomenal consciousness? We have no definitive answer to this question. However, we do have a conjecture as to at least part of the answer, as we will go on to describe below.

Merker's Evolutionary Pressure for Phenomenal Consciousness

The neurobiologist Bjorn Merker has suggested one plausible selection pressure that may have served to increase the evolutionary fitness of phenomenal consciousness in humans and other conscious animals. He points out that phenomenal consciousness produces a stable, coherent perceptual world for animals by distinguishing real motion in the world from apparent motion produced by the movement of sensory receptors (Merker 2005). One can experience the loss of this stable, coherent sensory world by a simple experiment. Close one eye and press gently with an index finger on the lower eyelid of the open eye. The movement of the eyeball produces an apparent motion of whatever is present in the experimenter's perceptual field. This external intervention therefore defeats the normal compensatory mechanisms that keep our subjective perceptual world stable. But when the constant movements of eyes, the head, and the body are endogenously controlled, no such movement of the world is perceivable. Thus, brain mechanisms underlying conscious perception must act to keep the world stable in spite of a vast and complex variety of movements in which we normally engage.

Merker does not claim that phenomenal consciousness is the only process capable of producing such a stable, coherent perceptual world. Nor does he claim that this process of distinguishing and suppressing apparent motion provides the only evolutionary selection pressure. He simply suggests that providing perceptual stability and coherency is one fitness benefit of phenomenal consciousness. But, what has all this to do with consciousness in machines?

A Perceptually Stable and Coherent LIDA Controlled Robot

In a commentary on Merker's article, Franklin suggested that producing a robot provided with a stable, coherent perceptual world might be a step toward a phenomenally conscious machine (2005). Let us call a sense organ *spatially sensitive* if movement of the organ produces apparent motion at its surface independent of what is happening in the environment. Any autonomous, mobile robot will likely require spatially sensitive sensory mechanisms, for example, vision, for moving appropriately in its world. Thus, the problem of distinguishing real motion from self-produced, apparent motion will be ubiquitous among such robots. One solution would be to build in mechanisms to shield the robot's action selection from apparent motion self-produced by its own movement of its sense organs. Such shielding mechanisms might conceivably be based on any of several different principles. One such principle would have the robot construct its own individual, coherent, and stable world, suppressing self-produced apparent motion,

as Merker argues that consciousness does for some animals. Such a stable, coherent perceptual world would prevent self-induced apparent motion from interfering with the robot's action selection.

Here we propose a LIDA controlled autonomous mobile robot with such a built-in shielding mechanism producing a coherent, stable, perceptual worldview. LIDA (Learning IDA) is a conceptual, and partially computational, cognitive architecture (Franklin and Patterson 2006), derived from IDA primarily by adding several modes of learning. The most accessible description of the LIDA architecture can be found on the web at http://ccrg.cs.memphis.edu/tutorial/index.html.

Can such a shielding mechanism be designed? That is an empirical question for robot designers. Our experience with designing IDA and LIDA suggests that essentially any human cognitive process, including deliberation and volitional decision making (Franklin 2000), can be effectively simulated in a software agent. Why not in a robot?

Is Phenomenal Consciousness Possible in such a Robot?

We humans attribute phenomenal consciousness to other humans because we experience it in ourselves, and perceive others as being similar to us. Most of us don't take seriously the possibility of a zombie, in the philosophical sense (Chalmers 1995), because there is no evidence that such a being could exist. Attribution of phenomenal consciousness to animals often results, as with humans, from the similarity of their nervous systems to ours (Seth, Baars, and Edelman 2005), or from the similarity of their behaviors to ours. But, why might one attribute phenomenal consciousness to a robot? Certainly not because of any similarity of nervous systems. Perhaps because of similarity in behavior. We would have no problem attributing phenomenal consciousness to a robot such as Star Trek the Next Generation's Commander Data, were he real rather than fictional. Recent experimental evidence suggests the likelihood of such attribution to artificial entities who behave like humans. Another possibility is attribution because of the similarity in the control architecture (mind) of the agent, be it human or robot.

Might a LIDA controlled robot that produces a stable, coherent perceptual world, as described above, be subjectively conscious? It would seem at least possible for several reasons. Such a robot would be functionally conscious. Based on the LIDA architecture, which is both psychologically and neuroscientifically grounded, its control structure would be quite similar to that of a human. In addition, it would satisfy Merker's coherent, stable perceptual world condition. But, might not other, additional, and as yet unknown, processes be needed in order to enable phenomenal consciousness in a robot? Indeed, they might. Note how Merker's work gives direction to robot designers attempting to produce conscious robots. We claim that building a robot as described above might well prove to be a significant step towards producing a phenomenally conscious robot.

References

Aleksander, I. 2000. *How to Build a Mind*. London: Weidenfeld and Nicolson.

Baars, Bernard J. 1988. *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Baars, BJ. 2003. How brain reveals mind: Neural studies support the fundamental role of conscious experience. *Journal of Consciousness Studies* 10:100-14.

Chalmers, David. 1995. Facing up to the problem of consciousness. *Journal of Consciousness Studies* 2:200-19.

Cotterill, RMJ. 2003. Cyberchild: A simulation test-bed for consciousness studies. *Journal of Consciousness Studies* 10:31-45.

Franklin, Stan. 2000. Deliberation and voluntary action in 'conscious' software agents. *Neural Network World* 10:505-21

Franklin, Stan. 2001. Automating human information agents. In *Practical applications of intelligent agents*, edited by Z Chen and LC Jain. 27-58. Berlin: Springer-Verlag.

Franklin, Stan. 2003. IDA: A conscious artifact? *Journal of Consciousness Studies* 10:47-66.

Franklin, Stan. 2005. Evolutionary pressures and a stable world for animals and robots: A commentary on Merker. *Consciousness and Cognition* 14:115-18.

Franklin, Stan and AC Graesser. 1997. Is it an agent, or just a program?: A taxonomy for autonomous agents. In *Intelligent agents iii*. 21-35. Berlin: Springer Verlag.

Franklin, S, A. Kelemen, and L. McCauley. 1998. IDA: A Cognitive Agent Architecture. In *IEEE Conf on Systems, Man and Cybernetics*. IEEE Press.

Franklin, Stan and FG Patterson Jr. 2006. The LIDA architecture: Adding new modes of learning to an intelligent, autonomous, software agent. In *IDPT-2006 Proceedings Integrated Design and Process Technology*: Society for Design and Process Science.

Haikonen, Pentti O. 2007. *Robot brains; circuits and systems for conscious machines*. UK: Wiley and Sons.

Holland, Owen. 2007. A strongly embodied approach to machine consciousness. *Journal of Consciousness Studies*. Special Issue on Machine Consciousness.

Krach, S, F Hegel, B Wrede, G Sagerer, and F Binkofski. 2008. Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE* 3:e2597.

Merker, Bjorn. 2005. The liabilities of mobility: A selection pressure for the transition to consciousness in animal evolution. *Consciousness and Cognition* 14:89-114.

Seth, AK, BJ Baars, and DB Edelman. 2005. Criteria for consciousness in humans and other mammals. *Consciousness and Cognition* 14:119-39.

Shanahan, MP. 2006. A cognitive architecture that combines internal simulation with a global workspace. *Consciousness and Cognition* 15:433-49.

More on Explaining a Gap

Gilbert Harman
Princeton University

In "Explaining an Explanatory Gap" (Harman 2007) I argued "that a purely objective account of conscious experience cannot always by itself give an understanding of what it is like to have that experience." Following Nagel (1974), I suggested that such a gap "has no obvious metaphysical implications. It [merely] reflects the distinction between two kinds of understanding," objective and subjective, where subjective understanding or "Das Verstehen" (Dilthey 1883/1989) of another creature's experience involves knowing what it is like to have that experience—knowing what sort of experience of one's own would correspond to the other creature's experience.

In the linguistic case, one understands one's own words in the sense that one is "at home" with them. One best understands what others say by translation into one's own way of using language. So it seemed to me useful to think of understanding another's experiences as a kind of translation into one's own. I suggested that we might be able to narrow the explanatory gap via an objective account of translation, e.g., in terms of functional relations. Such an account could be used in order to discover what it is like for another creature to have a certain objectively described experience given the satisfaction of two requirements. "First, one must be able to identify one

objectively described conceptual system as one's own [an identification that is not itself fully objective]. Second, one must have in that system something with the same or similar functional properties as the given experience."

With this brief background I would like first to discuss three commentaries on my paper (Harman 2007) recently published in this *Newsletter*. I then want to say something about the featured article in this issue, by Franklin, Baars, and Ramamurthy, which raises a somewhat different issue about whether a machine could have "the subjective experience of qualia...phenomenal qualia..."

Ledwig

Ledwig (2007) points to an unclarity as to exactly what is involved in subjective understanding or Das Verstehen and my suggestion that a special case of such understanding is the understanding one has of one's own language. What is meant by "one's own language"? Might it be a language like English, which has a history and a number of variants or dialects? Or is it a particular idiolect or (as linguists say) I-language?

I was thinking of a language in this second sense, one's I-language, the language in which one is at home, whose principles are internalized and not usually known to one in any serious way. Language in the social and historical sense is varied, containing words with which one is not familiar or which are used in different ways from the way one uses them. A language in the social and historical sense is not something one could be fully at home in. To understand what someone else says one must be able to find a translation or equivalent expression in one's I-language, one's particular way of using language.

Consider Ledwig's worry that

> it is not obvious to me whether a proper *Verstehen* in Harman's view also involves knowing where a certain meaning has come from or not. Or is a proper *Verstehen* also reflected by knowing under what kind of conditions one uses an expression? As a native German who has English as a second language, I now know under what kind of conditions it is appropriate to use the word "gorgeous" in American English, but I still find it truly surprising and puzzling that...[it] just applies to one sex.

My response is that the sort of understanding I had in mind need not involve any explicit knowledge of the conditions under which expressions are or are not appropriately used. The typical speaker of American English need not have considered the question whether the word "gorgeous" appropriately applies only to one sex, for example. For such a speaker, the understanding of this word is internalized, second nature.

Ledwig observes that an explanation of this fact about how "gorgeous" is used in American English might properly receive "not only a historical explanation...but also a cultural one." She suggests that "a historical or cultural explanation of the term would have helped me to understand its usage fully."

My response is that there are two ways of understanding the use of a word, one in which one is at home with using the word, as she was not completely at home using "gorgeous," the other in which one has a more objective understanding of the use of the word, involving explicit knowledge that it tends to be appropriately used only of one sex, for example.

She asks whether there can be "partial *Verstehen* or whether *Verstehen* always has to be complete," taking her understanding of "gorgeous" "to suggest that partial *Verstehen* is possible." I agree. One's subjective understanding of someone else is often partial in this way.

Concerning my suggestion that the explanatory gap reflects the difference between objective and subjective understanding, Ledwig wonders "whether the explanatory gap is inevitable." She notes various ways in which we may be able to use objective empirical methods to help gain understanding of others. For example, in order to gain an understanding of what it is like to be a member of the opposite sex, one might take on the identity of someone of the opposite sex, "one can even have a sex-change operation." She discusses a number of other cases, including trying to help a congenitally blind person gain an understanding of what it is to experience color surfaces: "one could enhance the texture of these surfaces to the blind."

Her interesting discussion of these and other cases is quite suggestive, while it seems to me to support the idea that the explanatory gap arises from the difference between objective and subjective understanding.

Worley

Worley (2007) appears to disagree with this last discussion of Ledwig's, when she says, "some concepts (the objective, third personal ones) are available to anyone with appropriate exposure and acculturation, and others are not."

Worley adds, "It is the status of these peculiarly first personal phenomenal concepts that Nagel finds mysterious and the appeal to *Verstehen* and failure of translation does not help resolve this mystery." I am not sure I agree with this. What Nagel finds mysterious (if that's the right word) is how to relate the two sorts of concepts or ways of understanding things. My own suggestion, repeated in the next section, is that we might make some progress on resolving this mystery if we were able to arrive at an objective understanding of what it takes for good translation between different subjective outlooks.

Nagasawa

Nagasawa (2007) argues that "Harman's formulation of the explanatory gap seems therefore to face the following difficulty: Either (i) it is irrelevant to the cogency of physicalism or (ii) if it *is* relevant, any talk of translation is otiose." I agree with (i), indeed, I explicitly said that the explanatory gap "has no obvious metaphysical implications" and "reflects the distinction between two kinds of understanding." On the other hand, with respect to (ii), I do not think that talk of translation is completely otiose in this connection.

As Nagasawa points out, simply knowing what it is like to undergo a physical process objectively described does not by itself eliminate the explanatory gap, because we may still wonder *why* that objectively described process is associated with this subjective experience. Merely having a way to translate between a creature's experiences and one's own does not eliminate the gap because the creature might be oneself and the gap is still there in one's own case.

Now suppose one had in addition a completely objective account of "translation" from the possible experiences of a creature to those of another, an account in terms of objective functional relations, for example. And suppose in addition that one was able to identify a particular objectively described mental system as one's own. Then it seems to me one might have *some sort* of objective understanding of what it is like to have various experiences. I agree that this understanding need not be completely objective, since it would depend on being able to identify a particular system as one's own, which is not a purely objective matter. It is not a purely objective matter which creature is oneself.

Phenomenal Qualia

Franklin, Baars, and Ramamurthy suggest that it may be possible actually to build a machine capable of consciousness. I agree

that, if by appeal to similarities between one's own functioning and behavior and that of the machine, one may find that it's possible to translate between events in some machine and one's own experiences in a way that satisfies certain conditions, then one can attribute consciousness to that machine and can even know what it's like to be a creature with those events occurring.

On the other hand, as the authors note, the lack of such a translation for a given machine would not rule out consciousness—that there is something that it is like to the machine to be such that machine; it would only rule out our being able to know in the relevant way what that is like.

The authors take the question whether a machine can be conscious to be the question whether a machine can have "the subjective experience of qualia...phenomenal consciousness..." I assume that by "qualia" they mean certain experienced qualities of perceived *objects*. But some philosophers use the word "qualia" to refer to intrinsic qualities of the *experience*, intrinsic qualities of which one is allegedly aware in having the experience. I deny that one is aware of intrinsic qualities of conscious experience. I say that to think we are aware of such intrinsic qualities of experience is to confuse qualities of an *experience* with qualities of the *object* of that experience.

The object of a conscious experience is an "intentional object"—an apparent object that may not really exist, like the dagger that MacBeth sees before him, or the pink elephants a drunk sees, or the Fountain of Youth that Ponce de Leon was looking for. The fact that the Fountain of Youth does not exist does not entail that Ponce de Leon wasn't looking for it. Similarly for the drunk's pink elephants and MacBeth's dagger. The drunk's elephants are pink and MacBeth's dagger drips with blood, but the drunk's experience isn't pink and MacBeth's experience isn't dripping with blood. It is a fallacy (the sense-datum fallacy) to suppose that features of the intentional object of experience are features of the experience. [Harman (1990/1999) discusses this further. Block (2007) argues the other side.] I am not saying that Franklin et al. commit this fallacy; only that some philosophers who use the "qualia" terminology do so.

Finally, I do not understand the authors' suggestion "that providing perceptual stability and coherency is one fitness benefit of phenomenal consciousness." I understand how providing perceptual stability and coherency benefits fitness. But I do not understand how this is provided by phenomenal consciousness, no matter how that is interpreted. Maybe what the authors mean merely is that perceptual stability is a feature of our perceptual consciousness, so a creature's having perceptual stability makes the creature more like us and so more like a conscious being. To that I agree.

References

Block, N. 2007. Consciousness, Accessibility and the Mesh between Psychology and Neuroscience. *Behavioral and Brain Sciences* 30:481-548.

Dilthey, W. 1883/1989. *Introduction to the Human Sciences*, edited by R. Makkreel and F. Rodi. Princeton, NJ: Princeton University Press. (Original work published 1883.)

Franklin, S. et al. 2008. A Phenomenally Conscious Robot? *American Philosophical Association Newsletter on Philosophy and Computers* 08(1).

Harman, G. 1990/1999. *The Intrinsic Quality of Experience. Philosophical Perspectives* 4:31-52; reprinted in Harman (1999), pp. 244-61.

Harman, G. 1999. *Reasoning, Meaning, and Mind*. Oxford: Clarendon Press.

Harman, G. 2007. Explaining an Explanatory Gap. *American Philosophical Association Newsletter on Philosophy and Computers* 06(2):2-3.

Ledwig, M. 2007. To Understand the Understanding—das Verstehen zu Verstehen: A Discussion of Harman's 'Explaining an Explanatory Gap'. *American Philosophical Association Newsletter on Philosophy and Computers* 07(1):16-18.

Nagel, T. 1974. What Is It Like to Be a Bat? *Philosophical Review* 83:435-50.

Nagasawa, Y. 2007. Formulating the Explanatory Gap. *American Philosophical Association Newsletter on Philosophy and Computers* 07(1):15-16.

Worley, S. 2007. Verstehen and the Explanatory Gap. *American Philosophical Association Newsletter on Philosophy and Computers* 07(2):15-16.

Quod Erat Demonstrandum.

Bernard J. Baars
The Neuroscience Institute, San Diego

Stan Franklin
University of Memphis

Uma Ramamurthy
St. Jude Children's Research Hospital

We welcome Professor Harman's commentary on our featured article in this issue of the *APA Newsletter*. We are pleased to be in agreement with his claim that "a purely objective account of conscious experience cannot always by itself give an understanding what it is like to have that experience" (Harmann 2007). Nevertheless, we can still make genuine progress on basic research goals like understanding conscious brains.

In past decades it has been unusual even to discover common terms of reference among scientists and philosophers interested in consciousness. But today we may be seeing a genuine convergence between our two disciplines. That is obviously a welcome development. We believe a phenomenally conscious robot could possibly be constructed, and that we are making measurable progress toward that goal. To build such a robot is not the same as sharing its subjective experiences. Nevertheless, given a clear set of empirical criteria that characterize conscious brains, we can demonstrate that the LIDA-Global Workspace approach is able to meet a growing subset of such criteria (Baars 1988; Seth, Baars, and Edelman 2005). That makes consciousness a natural-science problem like many others, without principled "explanatory gaps" that place limits on a steady growth in our understanding. If such metaphysical barriers exist, we have not yet encountered them.

Professor Harman makes a distinction between intrinsic subjective qualia, and subjectively inferred features of qualia. He writes:

> The authors take the question whether a machine can be conscious to be the question whether a machine can have "the subjective experience of qualia...phenomenal consciousness..." I assume that by "qualia" they mean certain experienced qualities of perceived objects. But some philosophers use the word "qualia" to refer to intrinsic qualities of the experience, intrinsic qualities of which one is allegedly aware in having the experience. I deny that one is aware of intrinsic qualities of conscious experience. I say that to think we are aware of such intrinsic qualities of experience is to confuse qualities of an experience with qualities of the object of that experience.

Psychologically, we believe that both of these claims may be true for adequate models of conscious perception. It is well-known in the scientific study of visual perception that human subjects are quite capable of reporting the retinal extent of a

visual object, such as the retinal projection of the sight of a book situated on a table in front of the subject. Thus, a subject can tell us that the book appears to span fifteen degrees of visual angle at a distance of about three feet. That is something that visual researchers do all the time in their own work, just as carpenters, painters, and architects have learned to do it. Such an estimate is clearly a subjective interpretation of the subjective sight of the book.

However, in addition, people can also estimate the veridical size of a book in the world of public objects. Veridical estimates are also subjective judgments, but through extensive practice, we can learn to make them quite accurately.

Thus the distinction Professor Harman makes appears to involve subjectivity in both aspects: A subjective estimate of a veridical or "public" aspect of a perceptual stimulus, and *also* a subjective estimate of the private extent of the visual stimulus—"what it looks like to an observer." Human beings can switch from public to private perspectives toward a host of stimulus dimensions, including colors, spatial dimensions, the effects of lighting, and even emotional or esthetic qualities.

These distinctions are well established in the research literature on visual perception. The sight of a book is not a single, conscious experience, but rather a sizable collection of subjective experiences, some of which are under voluntary control, and some of which allow us to make judgments about the veridical, public nature of the objects we experience.

Again, we are grateful for the opportunity to exchange views with a prominent philosopher, and we feel encouraged by the sense of improved communication on both sides. We may not be quite ready to write "QED"—"Which Was To Be Demonstrated"—under our research efforts, but we are all clearly moving in the right direction.

References

Baars, B.J. 1988. *A Cognitive Theory of Consciousness*. NY: Cambridge University Press.

Seth, A.K., Baars, B.J., and Edelman, D.B. 2005. Criteria for consciousness in humans and other mammals. *Consciousness and Cognition* 14:119-39.

Perception Loop and Machine Consciousness

Antonio Chella
Università di Palermo

Introduction

The current generation of systems for man-machine interaction shows impressive performances with respect to the external shapes, the mechanics, and the control of movements; see, for example, the *Geminoid* android robot developed by Ishiguro and colleagues.[1] However, these robots, currently at the state of the art, present only limited capabilities of perception, reasoning, and action in novel and unstructured environments. Moreover, the capabilities of user-robot interaction are standardized and tightly defined.

A new generation of robotic agents, able to perceive and act in unknown, dynamic, and unstructured environments should be able to pay attention to the relevant entities in the environment, to choose their own goals and motivations, and to decide how to reach them. To reach this result, a robotic agent must be able to simulate different functions of the human brain that allow humans to be aware of the environment that surrounds them, i.e., a robotic agent should show some form of machine consciousness.

Epigenetic robotics and synthetic approaches to robotics based on psychological and biological models have elicited many of the differences between the machine and mental studies of consciousness, while the importance of the interaction between the brain, the body, and the surrounding environment has been pointed out.

This paper takes into account the *externalist* (Rowlands 2003; Rockwell 2005) point of view by hypothesizing that the perception process is based on a *generalized loop* between the brain, body, and environment. The perception loop is in part internal and in part external to the robot, and it comprises the interactions among the proprioceptive and perceptive sensor data, the anticipations about the perceived scene, and the scene itself, through a focus of attention mechanism.

The perception model has been tested on an effective robot architecture implemented on an operating autonomous robot *ActivMedia PeopleBot* offering guided tours at the Archaeological Museum of Agrigento. Several public demos, some in the presence of media, validating the capabilities of the robot have been given over the last years.

A technical description of the perception loop is reported in Chella 2007. Here, the principles related with the machine consciousness debate are presented (see also Aleksander 2008).

Theoretical Remarks

Suggestions on how to implement a machine consciousness model based on externalism have been proposed in the literature. The most relevant one is due to O'Regan and Noë (O'Regan and Noë 2001), who discuss the process of visual awareness as based on *sensorimotor contingencies*. Following this approach, the robot should be equipped by a pool of sensorimotor contingencies so that entities in the environment activate the related contingencies that define the interaction schemas between the robot and the entity itself.

Some contingencies may be pre-programmed in the robot system by design (*phylogenetic contingencies*), but during the working life, the robot may acquire novel contingencies and therefore novel way of interacting with the environment. Moreover, the robot may acquire new ways of *mastery*, i.e., new ways to use and combine contingencies, in order to generate its own goal tasks and motivations (*ontogenetic* contingencies).

A mathematical analysis of the theory based on a simple robot in a simulated environment is presented in Philipona et al. (2003). The relationships between the sensorimotor contingencies and minimal axioms for consciousness has been analyzed in Aleksander and Morton (2005).

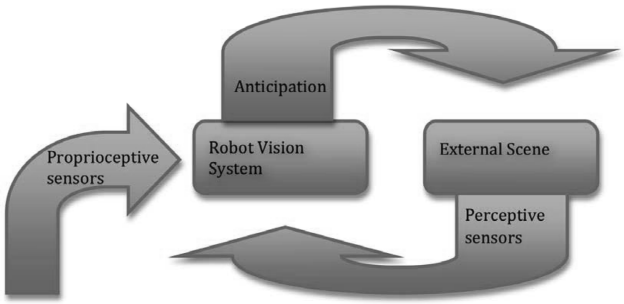
Manzotti and Tagliasco (Manzotti and Tagliasco 2005; Manzotti 2006 and 2007) proposed an externalist theory covering the phenomenal and the functional aspects of consciousness. They analyzed in detail the process that unifies the activity in the brain and the perceived events in the external world. Following this line, they propose some basic requirements for a "conscious" artefact.

The Robot Perception

The proposed robot perception model is inspired to the externalist approach following the lines of Manzotti and Tagliasco. The model is based on tight interactions between the robot brain, body, and environment. The model is sketched in Figure 1. *The Robot Vision System* receives in input the proprioceptive data from internal sensors as the odometric sensor, and the perceptive data from the external sensors, as the scene acquired by the video camera.

The perception loop works as follows: the robot vision system receives in input the robot position, speed, and so on

Fig. 1. The robot perception loop.



from the proprioceptive sensors and it generates the scene *anticipations*, i.e., the expectations about the perceived scene. The perception loop is then closed by the perceptive sensors that acquire the effective scene by means of the video camera.

The process of generation of scene anticipations is performed by a computer graphics simulator generating the expected image scene on the basis of the robot movements. The mapping between the anticipated and the perceived scene is achieved through a *focus of attention mechanism* implemented by means of suitable recurrent neural networks with internal states (Elman 1990). A sequential attentive mechanism is hypothesized that suitably scans the perceived scene and, according to the hypotheses generated on the basis of the anticipation mechanism, it predicts and detects the interesting events occurring in the scene. Hence, starting from the incoming information, such a mechanism generates expectations and it makes contexts in which hypotheses may be verified and, if necessary, adjusted.

The focus of attention mechanism selects the relevant aspects of the acquired scene by sequentially scanning the image from the perceptive sensors and comparing them in the generated anticipated scene. The attention mechanism is crucial in determining which portions of the acquired scene match with the generated anticipation scene: not all true (and possibly useless) matches are considered, but only those that are judged to be relevant on the basis of the attentive process.

The match of a certain part of the acquired scene with the anticipated one in a certain situation will elicit the anticipation of other parts of the same scene in the current situation. In this case, the mechanism seeks for the corresponding scene parts in the current anticipated scene. This type of anticipation is called *synchronic* because it refers to the same situation scene.

The recognition of certain scene parts could also elicit the anticipation of evolutions of the arrangements of parts in the scene; i.e., the mechanism generates the expectations for other scene parts in subsequent anticipated situation scenes. This anticipation is called *diachronic*, in the sense that it involves subsequent configurations of image scenes. It should be noted that diachronic anticipations can be related with a situation perceived as the precondition of an action, and the corresponding situation expected as the effect of the action itself. In this way diachronic anticipations can prefigure the situation resulting as the outcome of a robot action.

Two main sources of anticipation are taken into account. On the one side, anticipations are generated on the basis of the structural information stored in the robot by design. These kinds of anticipations are called *phylogenetic*. On the other side, anticipations could also be generated by a purely *Hebbian* association between situations learned during the robot operations. These kind of anticipations are called *ontogenetic*. Both modalities contribute to the robot perception process.

Ontogenetic anticipations are acquired by *online learning* and *offline learning*. During the normal robot operations, when something unexpected happens, i.e., when the generated anticipation image scene does not match the scene acquired by the perceptive sensors, the robot vision system learns to associate, by a *Hebbian* mechanism, the current image scene with the new anticipation image through the previously described attention mechanism. This is the *online* mode of anticipation learning.

In the *offline* anticipation learning, the proposed framework for conscious perception is employed to allow the robot to imagine future sequences of actions to generate and learn novel anticipations. In fact, the signal from perceptive sensors is related to the perception of a situation of the world out there. In this mode, the robot vision system freely generates anticipations of the perceptive sensors, i.e., it freely *imagines* possible evolutions of scenes, and, therefore, possible interactions of the robot with the external world, without referring to a current external scene. In this way, new anticipations or new combinations of anticipations may be found and learned offline by the robot itself through the synchronic and diachronic attention mechanisms.

The described perception loop, which is in part internal and in part external to the robot, constitutes the perception experience of the robot, i.e., what the robot perceives at a given instant. The generalized perceptual loop is the stage in which the two flows of information, i.e., the anticipations data generated by the robot and the perceptive data coming from the scene, coexist and compete for consistency. This kind of perceiving is an active process, since it is based on the generation of the robot anticipations and driven by the external flow of information. The robot acquires evidence for what it perceives, and at the same time it interprets visual information according to its anticipations.

The Robot at Work

The ideas sketched here have been implemented in *Cicerobot*, an autonomous robot equipped with sonar, laser rangefinder, and a video camera mounted on a pan tilt. The robot has been employed as a museum tour guide since 2005, operating at the Archaeological Museum of Agrigento, Italy, offering guided tours in the *Sala Giove* of the museum (see Macaluso and Chella 2007 for technical details).

The robot controller includes a behavior-based architecture (Arkin 1998) equipped with standard reactive behaviors as the static and dynamic obstacle avoidance, the search of free space, the path following, and so on.

The *phylogenetic* anticipations are programmed in the robot system by design and stored in the robot memory. They are related to the architectural entities in the museum scene. During the working life, the robot may acquire novel anticipations and therefore novel expectations and novel way of interacting with the museum environment, by means of *ontogenetic* anticipations. During a standard museum visit, the robot activates its own anticipations. In this case, the robot has a low degree of conscious perception. When something unexpected happens, for example, the presence of a new object in the museum, the robot arises its own degree of awareness and it copes with the situation by mastering suitable anticipations.

These unexpected situations generate a trace in the robot memory in order to allow the robot to generate new anticipations and/or new ways of combining anticipations. In this case, new trajectories of the focus of attention mechanism will be learned by the robot. This is an example of the *online* anticipation learning mode.

Therefore, the robot, by its interaction with the environment, is able to modify its own goals or to generate new ones. A new object in the museum will generate new expectations related with the object and the subsequent modifications of the expectations related with the standard museum tour. Moreover, as previously stated, in the *offline* anticipation learning mode, the robot freely generates and learns sequences of novel museum situations.

Conclusions

Clark and Grush (1999) introduce the notion of *Minimal Robust Representationalism*, i.e., the minimal internal representations with the following capabilities: the operative conception is non-trivial; the identification of internal states as representations does explanatory work; the identification is empirically possible; and the identified states figures in biological cognition. It should be noted that the proposed perception loop owns all the capabilities required by Clark and Grush.

The perception loop is related to the concept of *sensorimotor* contingencies proposed by O'Regan and Noe. The external environment and also the robot itself activate the anticipations by the attention mechanism that defines the interaction schemas between the robot and the environment. So, for example, a vase, a window, the visitors, will activate the related robot anticipations by means of suitable scans of the focus of attention. Therefore, in agreement with the approach of O'Regan and Noe, the robot phenomenology grows up from the mastery of contingencies at the basis of the task to execution of the robot.

A related approach is described by Grush (2004), based on the concept of *emulator* in the fields of motor control and visual perception. The basic cognitive architecture proposed by Grush is made up by a feedback loop connecting a controller, a plant to be controlled, and the emulator of the plant. The loop is *pseudo-closed* in the sense that the feedback signal is not generated by the plant, but by the emulator, which parallels the plant and it receives as input an efferent copy of the control signal sent to the plant. A more advanced architecture takes into account the basic schema of the Kalman filter. Comparing the Grush architecture with the described model, the anticipation generation process may be seen as a sort of visual emulator of the robot scene; anyway, the proposed model stresses the role of the focus of attention mechanism as the mapping process between the perceived and the anticipated image scenes.

The proposed model may be a good starting point to investigate *conscious perception* and its relationship with *cognitive* perception and with perception based on *stimulus-response* (see Boltuc and Boltuc 2007). An interesting point, in the line of Nagel (1974), is that a robot may have a different consciousness of the world than we humans may have, because it may be equipped with several perceptive and proprioceptive sensors which have no correspondences in human sensors, like, for example, the laser rangefinder, the odometer, the GPS, the WiFi or other radio links, and so on. Therefore, the line of investigation may lead to study new modes of conscious perception which may be alternative to human conscious perception, as, for example, the conscious perception of an intelligent environment, the conscious perception distributed in a network where the robots are network nodes, the conscious perception of a multirobot team, the robot with multiple parallel consciousness, and similar kinds of robot conscious perception.

Endnotes

1. Information is available online at http://www.irc.atr.jp/Geminoid/.

References

Aleksander, I. 2008. Machine consciousness. *Scholarpedia* 3(2):4162.

Aleksander, I. and Morton, H. 2005. Enacted Theories of Visual Awareness, A Neuromodelling Analysis. In *Proc. BVAI 2005*, LNCS 3704. 245-57. Heidelberg: Springer-Verlag.

Arkin, R.C. 1998. *Behavior-Based Robotics*. Cambridge, MA: MIT Press.

Boltuc, N. and Boltuc, P. 2007. Replication of the Hard Problem of Consciousness in AI and Bio-AI: An Early Conceptual Framework, paper from the 2007 AAAI Fall Symposium, edited by A. Chella and R. Manzotti. Technical Report FS-07-01. Association for the Advancement of Artificial Intelligence, Menlo Park, California.

Chella, A. 2007. Towards Robot Conscious Perception. In *Artificial Consciousness*, edited by A. Chella and R. Manzotti. 125-40. Imprint Academic, Exeter UK.

Clark, A. and Grush, R. 1999. Towards a Cognitive Robotics. *Adaptive Behavior* 7:5-16.

Elman, J.L. 1990. Finding Structure in Time. *Cognitive Science* 14:179-211.

Grush, R. 2004. The emulator theory of representation: motor control, imagery and perception. *Behavioral and Brain Sciences* 27:377-442.

Macaluso, I, and Chella, A. 2007. Machine Consciousness in CiceRobot, a Museum Guide Robot, paper from the 2007 AAAI Fall Symposium, edited by A. Chella and R. Manzotti. Technical Report FS-07-01. Association for the Advancement of Artificial Intelligence, Menlo Park, California.

Manzotti, R. 2006. A process oriented view of conscious perception. *Journal of Consciousness Studies* 13(6):7-41.

———. 2007. Towards artificial consciousness. *APA Newsletter on Philosophy and Computers* 07(1):12-15.

Manzotti, R. and Tagliasco, V. 2005. From behaviour-based robots to motivation-based robots. *Robotics and Autonomous Systems* 51:175-90.

Nagel, T. 1974. What is it like to be a bat? *Philosophical Revue* 83:435-50.

O'Regan, J.K. and Noë, A. 2001. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24:939-1031.

Philipona, D., O'Regan, J.K., and Nadal, J.P. 2003. Is There Something Out There? Inferring Space from Sensorimotor Dependencies. *Neural Computation* 15:2029-49.

Rockwell, W.T. 2005. *Neither Brain nor Ghost*. Cambridge, MA: MIT Press.

Rowlands, M. 2003. *Externalism – Putting Mind and World Back Together Again*. Montreal & Kingston: McGill-Queen's University Press.

The Fourth Way: A Comment on Halpin's "Philosophical Engineering"

Michael Wheeler
University of Stirling

It is common these days to distinguish between three kinds of cognitive science or artificial intelligence: classical, connectionist, and (something like) embodied-embedded. Of course, all such attempts at neat-and-tidy categorization are undoubtedly guilty of over-simplification in one way or another. For example, researchers sometimes build models that combine aspects of more than one approach (e.g., when conventional connectionist networks are used as control systems for embodied agents). That noted, however, one method for separating out our three kinds of cognitive science, so as to understand more clearly their basic theoretical commitments, would be to identify, in a very general way, the sorts of machine that each takes to capture the fundamental character of intelligence. If we adopt this strategy, classicism will be defined by the manipulation of symbols using structure-sensitive processes, connectionism by unfolding patterns of

activity in neurally inspired networks of simple processing units, and embodied-embedded thinking by complete autonomous robots engaged in perceptually guided motor activity. One of the many fascinating claims in Harry Halpin's strikingly original article "Philosophical Engineering: Towards a Philosophy of the Web" (2008) is that the Web constitutes a fourth conceptual anchor for the notion of mind as machine. Halpin's view, in short, is that the Web provides a general model of a computational machine that compels us to rethink the notion of representation, while simultaneously radicalizing our conception of cognition through a vindication of the idea that minds may be realized partly by factors located beyond the skin. In this comment on Halpin's article, I shall engage briefly with just some of the issues that confront us once we take this fourth way.

With apologies for the immediate whiff of self-centerdness, I shall begin by considering an argument from Halpin's paper that responds explicitly to some of my own previous work. I have been known to claim (e.g., Wheeler 2005) that any adequate account of representational explanation in cognitive science must have the consequence that while certain inner (within-the-skin) elements count as representations, most external (beyond-the-skin) elements don't. The justification for this restriction is largely methodological: it seems likely that *neural* states and processes do something that is, for the most part, psychologically distinctive, and we expect the concept of representation to help us explain how that something comes about. Thus, the constraint at issue may be specified more carefully as what I call the *neural assumption*. The neural assumption states that if intelligent action is to be explained in representational terms, then whatever criteria are proposed as sufficient conditions for representation-hood, they should not be satisfied by any extra-neural elements for which it would be unreasonable, extravagant, or explanatorily inefficacious to claim that the contribution to intelligent action made by those elements is representational in character. For if such illegitimate external factors qualified as representations, the claim that some neural state has a representational character would fail to single out what was special about the causal contribution to intelligent behavior made by that state. Notice that the neural assumption, as formulated, is liberal enough to allow *some* external factors to qualify as representations in the sense that is relevant for cognitive-scientific explanation. However, it is clear that representations construed this way will remain largely inside the head.

Halpin distances himself from this approach to representation, arguing that once our intellectual goal becomes a philosophy of the Web, as opposed to a philosophical account of how representation figures as an explanatory primitive in cognitive science, any inner-focused account of representation (such as my own) will fail to deliver what theory demands. As he puts it, the Web is "nothing if not a robustly representational system, and a large amount of research on the Web focuses on how to enable increasingly powerful and flexible forms of representations" (Halpin 2008, 6). Thus, "[w]hat we need is a notion of what a representation is, a definition that applies to both 'internal' and 'external' representations, not conditions for a representational explanation in cognitive science" (Halpin 2008, 7). In other words, what a philosophy of the Web requires is a suitably generic, locationally uncommitted account of representation that, in principle, applies equally to internal representations (those located within the skin) and external representations (those located outside the skin, such as those on the Web). Without such an account, we will be unable to make sense of the Web as a representational system. In the light of this analysis, Halpin proceeds to sketch a proposal for what it is for something to be a representation. Here he draws, in part, on Smith's (1996) notion of representation via *registration*, according to which the distinction between subject and object, and thereby between representation and represented, emerges from the dynamics of certain physical processes in which one region of space-time tracks the behavior of another.

I will be concerned not with the plausibility of Halpin's positive proposal, but rather with the alleged need for any unitary, locationally uncommitted account of representation. For it seems to me that, from the present perspective, although we need *a* concept of representation that illuminates the character of representational explanation in cognitive science, plus *a* concept of representation that makes sense of external representations (and thereby of the Web as a representational system), there is no reason to think that it must be the *same* concept of representation in both cases. Indeed, there are considerations which suggest that theoretically significant differences are to be expected. For example, when external representations are used to guide intelligent behavior, they do so via perception-action loops. Thus, consider familiar cases of visual maps, whether paper or electronic. Such representations are able to direct behavior because the agent *looks at* and *performs an embodied spatial manipulation of* the map-realizing elements (the atlas or the PDA). No such perception-action engagement with the behavior-guiding representations is present when we use neurally realized internal maps (assuming there are such things) to navigate around the world. One might expect these sorts of differences to have an impact upon the nature of the representations in question. Moreover, Halpin himself identifies certain principles that (he argues) not only characterize the external representations used by the Web, but also perhaps explain the intelligence-facilitating effects of the Web. It is hard to see how these principles (universality, inconsistency, self-description, least power, and the open world – see Halpin 2008, 9, for the details) apply to neural representations.

Of course, given this pattern of divergence, we need some reason to conclude that what we call internal representations and what we call external representations are both genuine members of some overarching category of representational elements. For this, however, it is sufficient that (a) the alternative notions be linked by the vague pre-theoretical thought that, to be a representation, a state or process should play some sort of standing-in-for function, and (b) there should be some sort of family resemblance structure in play. Evidence for (b) may be found in the observation that familiar cases of external representations (e.g., mathematical symbols) plausibly share certain properties with neural representations, properties such as multiple realizability and being the bearers of consumed information.

That said, Halpin's nervousness about my inner-focused account of representational explanation in cognitive science may have an alternative source. To see this, we need to plug in the relationship that, according to Halpin, exists between the external representationalism of the Web and (what is sometimes called) the *extended mind hypothesis* (Clark and Chalmers 1998). In general terms, those who believe in the extended mind hold that there are conditions under which cognitive states, processes, mechanisms, architecture, and so on may be partly realized by material elements located beyond the skin. Halpin's view is that the universal information space of the Web supplies a dynamic and open-ended suite of such elements. In other words, the ways in which we store, retrieve, manipulate, and transform representational structures on the Web mean that, under certain conditions, some of our cognitive traits are partly realized by those structures. Dramatic examples of such cognitive extension occur when multiple agents remotely access and update a shared map on the Web. In such cases, the "active manipulation of a representation

lets [the two agents] partially share a dynamic cognitive state and collaborate for their greater collective success…via shared external representations that are universally accessible over the Web" (Halpin 2008, 8).

This changes things. If we are to make sense of the Web not only as a representational system, but as a representational system whose elements may sometimes constitute part of an agent's cognitive architecture, then one might think that the pressure in the direction of a unitary account of representation increases. After all, given that certain representational structures on the Web are now to be granted cognitive status, it seems that an adequate account of representational explanation *in cognitive science* will need to apply not only to familiar inner elements such as neural states and processes, but also to those external structures. In other words, any purely inner-focused account of representation is now revealed as failing to deliver what *cognitive* theory demands. Although, to my mind, Halpin himself does not clearly separate out the present argument from the one with which we began (which doesn't turn on the putatively cognitive status of the external representational structures), it seems to be the present argument that offers the more compelling case for the view that we need a unitary, locationally uncommitted account of representation.

Halpin's analysis alerts us to the fact that once the idea of cognitive extension is on the table, the neural assumption (see above) needs to be separated out from what we might now dub the *global adequacy requirement*—the demand that we develop an account of representation suitable for the task of cognitive-scientific explanation. The latter is what Halpin (2008, 7) calls the "conditions for a representational explanation in cognitive science." In my previous work I have been guilty of running together the global adequacy requirement and the neural assumption (as indicated by the discussion of the neural assumption included above). Once we pull these analytical structures apart, however, we can see that the strategy of appealing to different notions of representation—the strategy that, as we saw, made sense of external representation under a non-cognitive interpretation—will also make sense of external representation under a cognitive interpretation. To see why this is, notice first that, depending on how one carves up nature into cognitive and non-cognitive regions, an account of representation that meets the neural assumption may not meet the global adequacy requirement. Consider: if externally located representations on the Web figure as genuine parts of cognitive processes, the global adequacy requirement will not be met by an account of representation that respects the neural assumption—or at least not by that account *on its own*. But that, of course, is the key point. For the global adequacy requirement may be met by a varied explanatory toolkit encompassing different notions of representation designed for different explanatory tasks, such as understanding how neural states contribute to intelligent behavior, and illuminating how external representations may figure as genuine parts of cognitive processes. What this suggests is that, with the extended mind added to the mix, and the concept of cognitive representational space expanded to include external structures, that space may reflect the same pattern of similarities and differences between internal and external representations that we identified earlier. Thus, even under a cognitive interpretation, the unitary notion of representation that Halpin seeks may be no more than a philosophical chimera.

So far, I have been assuming that Halpin is right that, under certain circumstances, the Web forms part of our cognitive resources. I now want to interrogate that idea—not, I hasten to add, because I think it's obviously wrong, but because we need to be clear about what a good argument for that conclusion would look like. The first thing to note here is that the extended mind hypothesis is a view about the whereabouts of mind that is distinct not only from the position adopted by orthodox cognitive science (classical or connectionist), but also from the position adopted by any merely embodied-embedded view. To illustrate this point, we can adapt an example originally due to Rumelhart et al. (1986). Most of us solve difficult multiplication problems using pen and paper. The pen and paper system is a beyond-the-skin factor that helps to transform a difficult cognitive problem into a set of simpler ones, and to temporarily store the results of intermediate calculations. For orthodox cognitive scientists and for supporters of the merely embodied-embedded view, that pen and paper system is to be conceived as a non-cognitive environmental prop. It is an external tool that aids certain cognitive processes via embodied interaction, but is not itself a proper part of those processes. Of course, orthodox cognitive scientists and embodied-embedded theorists differ on how best to characterize the interactive arrangement of skin-side cognitive processes and external prop. In particular, the embodied-embedded theorist is likely to count the bodily activity involved as itself a cognitive process, as opposed to a mere output of neurally located cognition, and to trace rather less of the source of the manifest complexity of the observed behavior to the brain, and rather more to the structured embodied interactions with the external pen and paper system. For all that, however, both of these camps think of cognition as a resolutely skin-side phenomenon. By contrast, the extended mind theorist considers the causally coupled combination of pen-and-paper resource, appropriate bodily manipulations, and in-the-head processing to be a cognitive system in its own right. We can now pinpoint the right question to ask: Does Halpin's analysis indicate that certain manipulations of the Web's universal information space constitute genuine cases of cognitive extension rather than merely embodied-embedded intelligence?

Halpin sometimes seem to suggest that cognitive extension results whenever an adaptive causal coupling between inner and outer elements produces an intelligent outcome. Thus, recall his example of two agents whose intelligent behavior is structured by shared remote access, via mobile telephones, to a Web page containing a map. He implies that coupling considerations are sufficient for cognitive extension when he writes that "[s]ince [the two agents] are sharing the representation and their behavior is normatively successful based on its use, [they] can be said to partially share the same cognitive state" (Halpin 2008, 8). A more sophisticated version of the coupling argument for cognitive extension emerges during Halpin's subsequent discussion of the ways in which the coupled combination of analogue organic processing with external digital computer memory enable human beings to succeed at cognitive tasks that are poorly tackled by unaided organic processing. This is a particularly striking example of the ways in which human cognition may be transformed through the integration of internal processing with external props and scaffolds that possess a different range of fundamental properties. Unfortunately, however, even given the transformative effects brought about by integrated bio-technological couplings, we don't yet have an argument for cognitive extension. As Adams and Aizawa (2008) forcefully point out, all coupling-based arguments for cognitive extension are dangerously insensitive to a crucial causal-constitutive distinction, that is, to the distinction between cognition being merely causally dependent on some factor, and to cognition being constituted by, or partly constituted by, that factor. The cognitive activities of Halpin's remote-map-using agents, as well as those of his digitally embedded brains, are surely causally dependent on external factors in ways to which traditional theorizing in cognitive science has been largely oblivious, but that is not enough to secure the cognitive status of those factors.

The main alternative to coupling-based arguments for cognitive extension is what, in the literature, is known as the *parity principle* (Clark and Chalmers 1998). Exactly how one should formulate the parity principle remains a matter of some dispute (Clark 2007; Wheeler forthcoming), but the general idea is that if there is functional equality with respect to governing behavior, between the causal contribution of certain internal elements and the causal contribution of certain external elements, and if the internal elements concerned qualify as the proper parts of a cognitive process, then there is no good reason to deny equivalent status—that is, cognitive status—to the relevant external elements. Halpin (2008, 8) quotes Clark and Chalmers' original statement of the parity principle, but it is unclear to what extent he gives weight to parity considerations as opposed to issues of coupling and integration. However, because the parity principle appeals, at root, to the notion of functional equivalence and not mere coupling, it does not run roughshod over the causal-constitutive distinction. So, provisionally at least, the parity-driven case for cognitive extension is on the firmer footing. In relation to Halpin's arguments, this prompts the following question: Is it ever correct to say that there is functional parity between (i) the causal contributions to intelligent behavior made by those inner factors that qualify as cognitive, and (ii) the causal contributions to intelligent behavior made by structures on the Web?

As far as I can tell, the answer to this question depends on the specific criteria that one thinks need to be satisfied for a causal contribution to count as cognitive, what Adams and Aizawa (2008) call the *mark of the cognitive*. Such criteria are necessary because, in order to deploy the parity principle, one must be able to isolate just those functions that inner elements perform that mark out their contribution as cognitive (e.g., the functions involved in the context sensitive storage and retrieval of information that might plausibly define the cognitive trait of memory). It is parity with respect to the realization of these particular functional roles that will establish the cognitive status of certain external elements. This introduces a complex issue that certainly cannot be settled here. It is worth noting, however, that if the extended mind theorist adopts a weak or promiscuous enough mark of the cognitive, then it will be easy enough to secure the result that cognition is extended; but the price of this success will be to welcome into the domain of the cognitive all kinds of wildly unlikely cases in a manner that ultimately casts doubt on the ability of the proposed mark to latch onto only what might be thought of as the proper objects of cognitive science. What this aspect of Halpin's project still needs, it seems, is a mark of the cognitive that allows certain external representations on the Web (such as remotely accessible maps just as they guide online intelligent behavior) to count as cognitive, while denying that same status to "wildly unlikely cases" (such as books in a home library or standing mobile telephone access to an Internet search engine meaning that one might dispositionally believe everything on the Web). Put in a more generic way, the problem is to find a path between the dual dangers of a kind of disproportionate elitism (excluding from the domain of the cognitive certain genuinely cognitive traits, just because they happen to be externally located) and a kind of excessive liberality (welcoming in to the domain of the cognitive certain unwanted interlopers, as a side-effect of making conceptual room for extended cognition). Halpin is not alone in facing this problem. Extended mind theorists in general have perhaps failed to realize just how much hangs on it. Nevertheless, it is a problem for Halpin, and one that, I think, he cannot ignore.

My response to Halpin's arguments has necessarily been selective. I could have written another comment purely on the issues that Halpin explores towards the end of his discussion, when he turns his attention to the relationship between bio-technological intelligence and the specific case of the Semantic Web. What I hope to have made manifest, however, is the rich vein of thought that runs through Halpin's paper. For while the power of the Web as a technological innovation is now beyond doubt, the potential power of the Web to have a conceptual impact on cognitive science remains under-appreciated. The second of these contributions is what I have called the fourth way, an intellectual path innovatively revealed by Halpin's article. My critical comments here do no more than point to twists and turns that, in my view, remain to be navigated as we explore that trail. The fourth way may well be the next way.

References

Adams, F. and Aizawa, K. 2008. *The Bounds of Cognition*. Malden, MA and Oxford: Blackwell.

Clark A. 2007. Curing Cognitive Hiccups: a Defense of the Extended Mind. *Journal of Philosophy* 104:163-92.

Clark, A. and Chalmers, D. 1998. The Extended Mind. *Analysis* 58(1):7-19.

Halpin, H. 2008. Philosophical Engineering: Towards a Philosophy of the Web. *APA Newsletter on Philosophy and Computers* 07(2):5-11.

Rumelhart, D.E., Smolensky, P., McClelland, J.L. and Hinton, G. 1986. Schemata and Sequential Thought Processes in PDP models. In *Parallel Distributed Processing: Explorations In The Microstructure of Cognition, Vol. 2: Psychological and Biological Models*, edited by J.L. McClelland and D. Rumelhart. 7-57. Cambridge, Mass.: MIT Press.

Smith, B. 1996. *On the Origin of Objects*. Cambridge Mass.: MIT Press.

Wheeler, M. 2005. *Reconstructing The Cognitive World: The Next Step*. Cambridge, Mass.: MIT Press.

Wheeler, M. Forthcoming. Minds, Things, and Materiality. In *The Cognitive Life of Things: Recasting the Boundaries of the Mind*, edited by C. Renfrew and L. Malafouris. Cambridge: McDonald Institute for Archaeological Research Publications.

Discussion Articles on Floridi

Toward a Metaphysical Foundation for Information Ethics

Terrell Ward Bynum
Southern Connecticut State University

1. Introduction

In recent philosophical writings, the name "information ethics" has been used to refer to a broad new branch of philosophy which includes diverse subfields like computer ethics, Internet ethics, agent ethics, virtual reality ethics, genetic technology ethics, neurotechnology ethics, and even nanotechnology ethics. That same name has been used by Luciano Floridi to refer to a specific, rigorous "macroethics," which he developed to provide a foundation for computer ethics. (See, for example, his article "Understanding Information Ethics" in this *Newsletter*, Fall 2007.) To keep these two different meanings of the term "information ethics" separate, the present article employs 'information ethics' (using regular, lower-case letters) to refer to the broad new branch of philosophy, and 'INFORMATION ETHICS' (using SMALL CAPS) to refer to Floridi's "macroethics."

Today the broad field of information ethics has become important for understanding (1) new human relationships and communities, (2) the ethical development and control of emerging technologies, (3) the preservation and advancement of human values, and (4) the enhancement of respect and

cooperation among the many diverse cultures interacting on the World Wide Web (see Bynum 2006). To achieve these beneficial goals, a robust metaphysical foundation for information ethics would be very helpful, and so the present article explores (1) relevant metaphysical ideas introduced by Norbert Wiener in his information ethics writings, and (2) developments in contemporary science that lend support to Wiener's metaphysics. Floridi's INFORMATION ETHICS theory is based upon a very different—Spinozian and Platonistic—metaphysics; and that is briefly discussed near the end of the present article.

2. Metaphysical ideas underlying Wiener's information ethics works

Significant metaphysical presuppositions relevant to information ethics were used decades ago by Norbert Wiener in his books *Cybernetics* (1948), *The Human Use of Human Beings* (1950, 1954) and *God & Golem, Inc.* (1964). Wiener's assumptions about the ultimate nature of the universe included his view that *information is physical*—subject to the laws of nature and measurable by science. The sort of information that he had in mind is sometimes called "Shannon information"[1]—named for Claude Shannon, who had been a student and colleague of Wiener's. Shannon information is the sort that is carried in telephone wires, TV cables, and radio signals. It is the kind of information that digital computers process and DNA encodes within the cells of all living organisms. Wiener believed that such information, even though it is physical, *is neither matter nor energy*. Thus, while discussing thinking as information processing in the brain, he wrote that the brain

> does not secrete thought "as the liver does bile," as the earlier materialists claimed, nor does it put it out in the form of energy, as the muscle puts out its activity. Information is information, not matter or energy. No materialism which does not admit this can survive at the present day. (Wiener 1948, 155)

According to Wiener's metaphysics, matter-energy and Shannon information are different physical phenomena, but neither can exist without the other. So-called "physical objects"—including living organisms—are actually persisting patterns of Shannon information encoded within an ever-changing flux of matter-energy. Every physical process is a mixing and mingling of matter-energy with information—a creative "coming-to-be" and a destructive "fading away"—as old patterns of matter-energy-encoded information erode and new patterns emerge.

A related aspect of Wiener's metaphysics is his account of human nature and personal identity. Human beings, too, are *patterns of information that persist through changes in matter-energy*. Thus, in spite of continuous exchanges of matter-energy between a person's body and the world outside the body (via respiration, perspiration, excretion, and so on), the complex organization or *form* of a person—that is, *the pattern of information encoded within a person's body*—is maintained, thereby preserving life, functionality, and personal identity. As Wiener poetically said,

> We are but whirlpools in a river of ever-flowing water. We are not stuff that abides, but patterns that perpetuate themselves. (Wiener 1954, 96)
>
> . . .
>
> The individuality of the body is that of a flame…of a form rather than of a bit of substance. (Wiener 1954, 102)

To use today's language, humans are "information objects" whose personal identity is tied to bodily information processing and to persisting patterns of Shannon information within the body, rather than to specific bits of matter-energy that happen to make up one's body at any given time. Through breathing, drinking, eating, excreting, and other metabolic processes, the matter-energy that makes up one's body is constantly changing. Nevertheless, one remains the same person over time because the pattern of Shannon information encoded within the body remains essentially the same.

An additional aspect of Wiener's metaphysics is his account of *good and evil* within nature. He used the traditional distinction between "natural evil," caused by the forces of nature (for example, earthquakes, volcanoes, diseases, floods, tornados, and physical decay), and "moral evil" (for example, human-caused death, injury, pain, and sorrow). The ultimate natural evil, according to Wiener, is *entropy*—the loss of useful energy and useful Shannon information that occurs in virtually every physical change. According to the second law of thermodynamics, essentially all physical changes decrease available energy and available Shannon information. As a result, everything that ever comes into existence will decay and be destroyed. This includes any entity that a person might value, such as one's life, wealth, and happiness; great works of art; magnificent architectural structures; cities, cultures, and civilizations; the sun and moon and stars. None of these can survive the decay and destruction of entropy, for everything in the universe is subject to the second law of thermodynamics.

3. Recent developments in science that support Wiener's metaphysics

Wiener's intuitions in the 1940s about the ultimate nature of all entities in the universe, that *they consist of information encoded in matter-energy*, anticipated later research and discoveries in physics. During the past two decades, for example, physicists—starting with Princeton's John Wheeler (see Wheeler 1990)—have been developing a "theory of everything," which presupposes that the universe is fundamentally informational, that every physical "object" or entity is, in reality, a pattern or "flow" of information encoded in matter-energy. Wheeler's hypothesis has been studied and furthered by other scientists in recent years, and their findings support Wiener's metaphysical presuppositions. As explained by MIT professor Seth Lloyd:

> The universe is the biggest thing there is and the bit is the smallest possible chunk of information. The universe is made of bits. Every molecule, atom and elementary particle registers bits of information. Every interaction between those pieces of the universe processes that information by altering those bits. (Lloyd 2006, 3)
>
> . . .
>
> I suggest thinking about the world not simply as a machine, but as a *machine that processes information*. In this paradigm, there are two primary quantities, energy and information, standing on an equal footing and playing off each other. (Lloyd 2006, 169)

Science writer Charles Seife notes that "information is physical" and so,

> [Shannon] Information is not just an abstract concept, and it is not just facts or figures, dates or names. It is a concrete property of matter and energy that is quantifiable and measurable. It is every bit as real as the weight of a chunk of lead or the energy stored in an atomic warhead, and just like mass and energy, information is subject to a set of physical laws that

> dictate how it can behave—how information can be manipulated, transferred, duplicated, erased, or destroyed. And everything in the universe must obey the laws of information, because everything in the universe is shaped by the information it contains. (Seife 2006, 2)

In addition, the encoded Shannon information that constitutes every existing entity in the universe is *digital and finite*. Thus, Wheeler's one-time student Jacob Beckenstein discovered the so-called "Beckenstein bound," which is *the upper limit* of the amount of Shannon information that can be contained within a given volume of space. The maximum number of information units ("bits") that can fit into any volume is *fixed by the area of the boundary enclosing that space*—one bit per four "Planck squares"[2] of area (Beckenstein 2003). In summary, then, the information that constitutes all the existing entities in the universe is not infinite or infinitely divisible, it is instead finite and digital; and only so much information can be contained within a specific volume of space.

The metaphysical intuitions of Norbert Wiener, together with recent supportive developments in contemporary physics, provide a new account of the ultimate nature of the universe, a new understanding of life and of human nature, and even a new view—worthy of the "Information Age"—of all existing entitles in the universe: All are information objects or information processes. All living things, for example, store and process Shannon information in their genes and use that information to create the stuff of life, such as DNA, RNA, proteins, and amino acids. Nervous systems of animals take in, store, and process Shannon information making motion, perception, emotion, and thinking possible. And, as Charles Seife points out,

> Each creature on earth is a creature of information; information sits at the center of our cells, and information rattles around in our brains. ...Every particle in the universe, every electron, every atom, every particle not yet discovered, is packed with information…that can be transferred, processed, and dissipated. Each star in the universe, each one of the countless galaxies in the heavens, is packed full of information, information that can escape and travel. That information is always flowing, moving from place to place, spreading throughout the cosmos. (Seife 2006, 3)

Given Wiener's metaphysics, it follows that there are two physical phenomena, working together, which bring about and alter everything that exists. One is the transformation of matter-energy from one state to another, and the second is the "flow" (storage, preservation, alteration, transmission) of Shannon information, which constitutes every physical event and process. If these are, indeed, the two creative processes in the universe, they provide a powerful explanation for the world-transforming impacts of the Industrial Revolution in the nineteenth century and the Information Revolution of today.

The "heat engines" of the Industrial Revolution gave to human beings significant new control over one of the two creative processes in the universe: *matter-energy transformation*. The result was unprecedented power to alter the world in countless, and sometimes quite radical, ways.

Similarly, today's "information engines" (microchips, computers, information networks, cell phones, iPhones, iPods, computerized medical instruments, robots, digital weapons of war, and many, many more) are bestowing upon human beings dramatically increased control over the second creative process of the universe—*storage, manipulation, and transmission of Shannon information*. The Information Revolution, therefore, will surely change the world just as dramatically as the Industrial Revolution did a century ago. Indeed, even though our current "Information Age" has only just begun, already the unprecedented ability to store, shape, manipulate, and transmit Shannon information has resulted in a staggering number of social, political, economic, educational, medical, and military changes (to mention only a few example areas). And these changes have generated a seemingly endless number of ethical challenges involving, for example, privacy, security, ownership of intellectual property, preservation of human values, understanding among diverse cultures, ethical rules for robots and cyborgs, and so on (see Bynum 2008).

4. Casting new light on the history of computer ethics

The metaphysical ideas that Wiener employed in his information ethics writings cast new light upon important aspects of the history of computer ethics. Consider, for example, the success of Moor's classic definition of computer ethics (Moor 1985). The enormous new power that computing bestows upon human beings is its capacity to manipulate Shannon information in accordance with the laws of logic. Moor calls this ability "logical malleability," which makes computers "almost-universal tools" that can perform nearly any task. By generating so many new possibilities, computers and related information technologies create vast numbers of ethical "policy vacuums." That is, because of information technologies people can now do many things that they never could do before; and since no one could do those things before, there may be no laws or standards of good practice or other ethically relevant policies to govern them. Society needs to identify and ethically justify new policies to fill those vacuums. Wiener's metaphysical foundation for information ethics, therefore, explains why Moor's classic 1985 paper, regarding "logical malleability" and "policy vacuums," turned out to be such a persuasive and influential contribution to the field: If the "flow" of Shannon information is the second creative process of the universe, then the power to manipulate Shannon information, which is bestowed by information and communication technologies, is bound to make it possible for humans to do many new things for which ethical policies have not yet been established.

Wiener's metaphysical ideas also shed light upon another important development in the history of computer ethics, namely, the so-called "uniqueness debate," which began as a disagreement between Walter Maner and Deborah Johnson in the late 1970s. At that time, Maner coined the name "computer ethics" to refer to the new branch of applied ethics that he envisioned, one which was to be devoted to ethical problems "aggravated, altered, or even *created*" by computer technology. Johnson, who was on the same philosophy faculty with Maner, disagreed with his claim that computer technology creates *wholly new* ethical problems, although she did agree that computer technology can "give a new twist" to traditional ethical problems. Maner's 1978 *Starter Kit in Computer Ethics*, and Johnson's 1985 textbook *Computer Ethics* stated their initial assumptions about the possible uniqueness of problems in computer ethics. Later, in his ETHICOMP1995 keynote speech, "Unique Ethical Problems in Information Technology," Maner presented a strong defense of his "uniqueness" assumption (1996). This influential paper sparked a decade-long discussion, at computer ethics conferences and in computer ethics publications, that came to be known as the "uniqueness debate."[3] That debate included challenges such as this: *Why do computers need an ethics of their own? Other machines have had a big impact upon the world, but they don't have an ethics of their own. For example, there is no such thing as "sewing machine ethics" or "locomotive ethics" or "automobile*

ethics." So why should there be a "computer ethics"? (see Maner 1996).

Wiener's metaphysics provides an excellent answer to this, and similar, challenges: Computer technology gives human beings unprecedented control over one of the two fundamental creative processes of the universe. This explains the power of information technology to bring about the Information Revolution and thereby radically transform the world. Sewing machines, locomotives, and automobiles are three of the numerous components of the Industrial Revolution, but they do not account for the Industrial Revolution in the deep and profound sense in which computing technology accounts for the Information Revolution. Computers and related information technologies do indeed merit an ethics of their own!

5. The metaphysics of Floridian INFORMATION ETHICS

During the past decade, Luciano Floridi has developed a metaphysical foundation for his new "macroethics" theory, INFORMATION ETHICS. Upon first sight, Floridi's metaphysical foundation seems very similar to Wiener's metaphysical ideas, as well as to scientific findings of contemporary physics. Thus, according to Floridi, every existing entity in the universe, when viewed from a certain "level of abstraction," can be construed as an "informational object" with a characteristic data structure that constitutes its very nature. And, for this reason, Floridi says that the universe considered as a whole can be called "the infosphere." Each entity in the infosphere (that is, every existing being) can be significantly damaged or destroyed by altering its characteristic data structure and thereby preventing it from "flourishing." Such damage or destruction Floridi calls "entropy," which results in "empoverishment of the infosphere" (in other words, damage to the universe as a whole). Entropy therefore constitutes evil that should be avoided or minimized. With this in mind, Floridi offers four "fundamental principles" of INFORMATION ETHICS:

i. entropy ought not to be caused in the infosphere (null law)

ii. entropy ought to be prevented in the infosphere

iii. entropy ought to be removed from the infosphere

iv. the flourishing of informational entities as well as the whole infosphere ought to be promoted by preserving, cultivating, and enriching their properties

Although the metaphysics that underlies Floridi's INFORMATION ETHICS sounds much like Wiener's, it actually is very different. Wiener's theory, for example, is a form of *materialism grounded in the laws of physics*; while Floridi's theory presupposes *a Spinozian, even a Platonic, metaphysics* (Floridi 2007). In Floridi's INFORMATION ETHICS, but not in Wiener's metaphysics, non-living entities like databases, rivers, and stones have "rights" that ought to be respected. In addition, Floridian "entropy" is not the entropy of physics; and Floridian "information" is *not* Shannon information.[4] By construing every existing entity in the universe as an "informational object" that has at least a minimal moral worth, Floridi shifts the focus of ethical consideration away from the actions, characters, and values of human agents toward the "evil" (harm, dissolution, destruction)—"entropy"—suffered by objects in the infosphere. With this approach, every existing entity—humans, other animals, organizations, plants, even non-living artifacts, electronic objects in cyberspace, pieces of intellectual property, stones—can be interpreted as *potential agents* that affect other entities, and as *potential patients* that are affected by other entities. Thus, Floridi's INFORMATION ETHICS can be described as a "patient-based" non-anthropocentric ethical theory instead of the traditional "agent-based" anthropocentric ethical theories like utilitarianism, Kantianism, and Aristotelianism. And Floridi's underlying metaphysics is *not* the metaphysics of Wiener.

6. Concluding Remarks

The very broad field, which I have here called "information ethics," includes a number of sub-fields like computer ethics, Internet ethics, agent ethics, genetic technology ethics, neurotechnology ethics, and others. Because of the global reach of the Internet, and the resulting interaction of many different cultures around the world (see Gorniak 1996), information ethics, understood in this very broad sense, has become a major factor in the emergence of *a possible global ethics that can apply to peoples and cultures worldwide* (see Bynum 2006 and Tong 2008). Whether one adopts Wiener's materialist foundation for information ethics, or a Platonic Spinozian foundation similar to that of Floridi's INFORMATION ETHICS, or some yet-to-be-developed metaphysics, the implications for global ethics are likely to be significant as the Information Revolution dramatically changes the world.

Endnotes

1. Shannon information—the kind of information that is generated, processed, and stored by computers (and other ICTs)—is purely syntactical. It is physical in nature and obeys the laws of nature that physics studies. But Shannon information does not, in itself, have any meaning. Humans must assign semantic meaning to the syntactical structures of Shannon information. The same symbols and formulas within a computer, therefore, could represent any number of objects and processes, such as the size and trajectory of a missile, trends and features in a country's economy, germs and their actions within a human body, the positions and motions of a robotic arm, and so on and so on.

   Understanding exactly what semantic meaning is and how it is related to syntactical Shannon information is one of the major unresolved issues in the philosophy of information. Indeed, it is one of the most difficult problems in all of philosophy. Fortunately, the present paper does not have to assume any particular answer to this vexed question to achieve its goals.
2. A Planck square is a very, very tiny area: about 2.61223×10^{-70} square meters.
3. Available space does not permit a long discussion of the "uniqueness debate" here. For an extended discussion and references to the relevant literature, see Floridi and Sanders 2002.
4. For an in-depth discussion of Floridi's information ethics theory, see Floridi 2007.

References

Beckenstein, J.D. 2003. Information in the Holographic Universe. *Scientific American* August:58-65.

Bynum, T.W. 2006. Flourishing Ethics. *Ethics and Information Technology* 8(4):157-73.

Bynum, T.W. 2008. Computer and Information Ethics: Basic Concepts and Historical Overview. *Stanford Encyclopedia of Philosophy*. Accessed 7 July 2008 at http://plato.stanford.edu/

Floridi, L. 2007. Understanding Information Ethics *APA Newsletter on Philosophy and Computers* 07(1):3-14. Accessed 13 October 2007 at http://www.apaonline.org/publications/newsletters/Vol07n1/Computers/04.asp.

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

Gorniak-Kocikowska, K. 1996. The Computer Revolution and the Problem of Global Ethics. In *Global Information Ethics*, edited by Bynum, T.W. and Rogerson, S. 177-90. Opragen Publications. [Published as a special issue of the journal *Science and Engineering Ethics*.]

Johnson, D.G. 1985. *Computer Ethics*. Prentice-Hall.

Lloyd, S. 2006. *Programming the Universe*. Knopf.

Maner, W. 1978, 1980. *Starter Kit in Computer Ethics*. Helvetia Press and the National Information and Resource Center for Teaching Philosophy, 1980. [Originally self-published in 1978.]

Moor, J.H. 1985. What Is Computer Ethics? In *Computers and Ethics*, edited by Bynum, T.W. 266-75. Blackwell. [Published as the October 1985 issue of the journal Metaphilosophy.]

Moor, J.H. 2008. Why We Need Better Ethics for Emerging Technologies. In *Information Technology and Moral Philosophy*, edited by van den Hoven, J. and Weckert, J. 26-39. Cambridge University Press.

Seife, C. 2006. *Decoding the Universe: How the New Science of Information is Explaining Everything in the Cosmos, from Our Brains to Black Holes*. Viking, the Penguin Group.

Tong, J. 2008. Toward a Global Information Ethics: Some Confucian and Aristotelian Considerations. In *Living, Working and Learning Beyond Technology*, edited by Bynum, T.W., Calzarossa, M., De Lotto, I., and Rogerson, S. Proceedings of ETHICOMP2008, University of Pavia Press.

Wheeler, J. 1990. *Information, Physics, Quantum: The Search for Links*. Westview.

Wiener, N. 1948. *Cybernetics: or Control and Communication in the Animal and the Machine*. Technology Press.

Wiener, N. 1950, 1954. *The Human Use of Human Beings*. Houghton Mifflin, 1950. Second Revised Edition, Doubleday Anchor, 1954.

Wiener, N. 1964. *God & Golem, Inc.: A Comment on Certain Points Where Cybernetics Impinges on Religion*. MIT Press (USA) and Chapman & Hall (London).

Too Much Information: Questioning Information Ethics

John Barker
University of Illinois at Springfield

1. Introduction

In a number of recent publications,[1] Luciano Floridi has argued for an ethical framework, called Information Ethics, that assigns special moral value to information. Simply put, Information Ethics holds that all beings, even inanimate beings, have intrinsic moral worth, and that existence is a more fundamental moral value than more traditional values such as happiness and life. Correspondingly, the most fundamental moral *evil* in the world, on this account, is entropy—this is not the entropy of thermodynamics, but entropy understood as "any kind of *destruction*, *corruption*, *pollution*, and *depletion* of informational objects" (Floridi 2007, 9). Floridi regards this moral outlook as a natural extension of environmental ethics, in which non-human entities are treated as possessors of intrinsic moral worth and, more specifically, of *land ethics*, where the sphere of moral patients is further extended to include inanimate but naturally occurring objects. On Floridi's view, artifacts can also be moral patients, including even such "virtual" artifacts as computer programs and web pages.

In general, then, Floridi holds that *all objects* have some moral claim on us, even if some have a weaker claim than others; moreover, they have this moral worth intrinsically, and not because of any special interest we take in them. In this paper, I want to consider the motivation and viability of Information Ethics as a moral framework. While I will not reach any firm conclusions, I will note some potential obstacles to any such moral theory.

2. Background and Motivation

Before continuing, we need to clarify the notions of *object* and *information*, as Floridi uses those terms. Briefly, an informational entity is any sort of instantiated structure, any pattern that is realized concretely. In particular, information is not to be understood semantically. An informational object need not have any semantic value; it need not represent the world as being this way or that. Instead, information should simply be thought of as structure.

Now any object whatsoever may be regarded as a realization of some structure or other. Floridi realizes this, and indeed expands on it in a view he terms "Informational Structural Realism" (ISR).[2] ISR is a metaphysical account of the world that basically dispenses with substrates in favor of structures. On this view, the world should be regarded as a system of realized structures, but it is a mistake to ask what substrate the structures are ultimately realized in: it is structures "all the way down."

ISR is a fascinating thesis, but it will not be my purpose here to offer any further examination or critique of it. I mention it simply to show that when Floridi speaks of informational entities, he is really speaking of *arbitrary* entities. Information ethics is, in fact, a theory of arbitrary objects as moral patients. By casting it in terms of information, Floridi is stressing that the class of entities we should be concerned about, as moral patients or otherwise, is broader than the familiar concrete objects of our everyday experience. It should include any sort of instantiated information whatsoever, be it a person, a piece of furniture, or a "virtual" web-based object.

Now Floridi's central claim, that all entities have some (possibly very minimal) moral claim on us, while fascinating, certainly runs counter to most moral theories that have been proposed. It therefore seems reasonable to ask for some argument for it, or at least some motivation. The main rationale Floridi provides seems to be an argument from precedent. Before people started thinking systematically about ethics, they withheld the status of moral patient from all but the members of their own tribe or nation. Later, this status was extended to the whole of humanity. Many if not most people would now treat at least some non-human animals as moral patients, and some would ascribe moral worth to entire ecosystems and even to inanimate parts of nature. Thus, the history of ethical thinking is one of successively widening the sphere of our moral concern, and the logical end result of this process is to extend our moral concern to all of existence—or so Floridi argues.

However, as it stands this argument seems weak. True, there has been some historical tendency for moral theories to broaden the sphere of appropriate targets of moral concern. This tendency may continue indefinitely, until all of existence is encompassed. And then again, it may not. Here it is worth considering *why* at least some non-human animals are now generally considered moral patients. The main rationale, both historically and for most contemporary moral theorists, is that animals have a capacity for pleasure and suffering. It does not matter for my purposes whether this is the only or best rationale for extending moral consideration to animals. The point is that *some* rationale was needed; the mere precedent of extending moral consideration from smaller to larger groups of humans was not itself a sufficient reason to further extend it to animals. Likewise, if we are to extend the sphere of moral patients still further, we will need a specific reason to do so, not just precedent.

The most ardent supporters of animal rights have always been Utilitarians, and Utilitarianism justifies the inclusion of animals with a specific account of what constitutes a benefit or harm. Namely, benefit and harm are identified with pleasure and suffering, respectively. Once this identification is made, all it takes to show that a given being is a moral patient is to show that it can experience pleasure and pain. If Floridi were to give a specific account of what constitutes benefit or harm to an arbitrary entity, that would go some way toward providing a

rationale for Information Ethics. It is also desirable for another reason. Floridi's ethical account is "patient-oriented" (Floridi 2007, 8). This may or may not mean that it is a consequentialist theory; however, it seems fair to assume that in a patient-centered moral theory, an action's benefit or harm to moral patients plays a preeminent role in determining its rightness or wrongness. Thus, it seems desirable for such a theory to include a general account of benefit and harm. Moreover, such a theory can presumably be action-guiding only if it provides at least some such account.

Does Floridi offer such an account of benefit and harm? He does identify good and evil with "existence" and "entropy," respectively; but as I will argue below, it is not clear that this amounts to a general account of benefit and harm, either to individual entities or to the universe or "infosphere" at large. Now to some extent this omission is understandable, given the pioneering nature of Floridi's work. However, I will argue below that there are substantial obstacles in principle to providing any such account. The notion of an arbitrary object, I suspect, is simply too broad to support any substantive account of harm and benefit.

3. Information and Entropy

Let us first see why there is even a question about fundamental good and evil in Information Ethics. Floridi identifies existence as the fundamental positive moral value, and inexistence as the fundamental negative value. Thus, it might seem natural to suppose that an action is beneficial if it creates (informational) objects, and harmful if it destroys them, with the net benefit or harm identified with the net number of objects created or destroyed.

The trouble with this proposal is that given the broad conception of objects that we are working with, *every* act both creates and destroys objects. Since objects are simply instantiated patterns, there are indefinitely many objects present in any given physical substrate. Any physical change whatsoever involves a change in the set of instantiated patterns, thus creating and destroying informational objects simultaneously. Moreover, even if it were possible to count the number of informational objects in a given medium, such a count would ignore the fact that some beings have more inherent moral worth than others: this is fairly obvious in its own right, and Floridi himself insists on it, asserting that some moral patients have a strong claim on us while for others, the claim is "minimal" and "overridable" (Floridi 2007, 9).

Thus, if we are to take seriously the idea that "being" is the most fundamental good and "inexistence" or "entropy" is the most fundamental evil, we cannot calculate good or evil by simply counting objects. A natural idea, and one which is somewhat suggested by Floridi's term "entropy," is that fundamental moral value should be identified with some overall measure of the informational richness or complexity of a system. This would preserve the idea of being and nonbeing as fundamental moral values while avoiding the difficulties involved in the simple counting approach.

One of the best-developed accounts of non-semantic information is *statistical information theory*. This theory, developed by Claude Shannon in the 1940s,[3] has been used very successfully to describe the amount of information in a signal without describing the signal's semantic content (if any). Thus, it seems like a natural starting point for describing the overall complexity or richness of a system of informational objects.

Statistical information theory essentially identifies high information content with low probability. Specifically, the Shannon information content of an individual message M is defined to be $\log_2(1/p(M))$, where $p(M)$ is the probability that M occurs.[4] As a special case, consider a set of 2^n messages, each equally likely to occur; then each message will have a probability of 2^{-n}, and an information content of $\log_2(2^n) = n$ bits, exactly as one would expect. The interesting case occurs when the probability distribution is non-uniform; low probability events occur relatively rarely, and thus convey more information when they do occur.

As is well known, the definition of Shannon information content is formally almost identical to that of statistical entropy in physics. The entropy S of a given physical system is defined to be $S = k_B \ln \Omega$, where k_B is a constant (Boltzmann's constant) and Ω is the number of *microstates* corresponding to the system's *macrostate*. (A system's macrostate is simply its macroscopic configuration, abstracting away from microscopic details; the corresponding microstates are those microscopic configurations that would produce that macrostate.) Now for a given microstate q and corresponding macrostate Q, Ω is simply the probability that the system is in microstate q given that it is in macrostate Q. In other words, the entropy of a system is simply $k_B \ln (1/p_Q(q))$, where p_Q is a uniform probability distribution over the microstates in Q. Alternatively, if we posit a uniform probability distribution p over all possible microstates q, then we have $p_Q(q) = p(q) / p(Q)$, and thus $S = (k_B/p(q)) \ln p(Q) = -(k_B/p(q)) \ln (1/p(Q))$; the quantity $k_B/p(q)$ is a constant because the measure p is uniform. In any case, we have $S = K \log (1/p)$, where K is a constant and p is the probability of the state in question under some probability measure (the base may be omitted on the log because it only affects the result up to a constant, and may thus be subsumed in K). Thus, up to a proportionality constant, statistical entropy is a special case of Shannon information content.

However, it is the *wrong* special case, since, as Floridi states very clearly, the fundamental evil which he refers to as "entropy" is not thermodynamic entropy. And, indeed, in light of the second law of thermodynamics, thermodynamic entropy is not a reasonable quantity for moral agents to try to minimize. Thus, if we are to use Shannon information theory to capture the morally relevant notion of complexity, we will have to use a probability measure other than that described above. However, information theory does not offer us any guidance here, because it does not specify a probability measure: it simply *assumes* some measure as given. Typically, when applying information theory, we are working with a family of messages with well-defined statistics; thus, a suitable p is supplied by the context of the problem at hand.

Thus, Shannon information theory provides a measure of a system's information content, but this measure is relative to a probability measure p. This presents an obstacle to explaining complexity in terms of Shannon information and simultaneously claiming that complexity is a fundamental, intrinsic moral value. If we allow complexity to be relative to a probability measure, then intrinsic moral worth will also be relative to a probability measure. Conceivably, different probability measures could yield wildly different measures of complexity and, thus, of intrinsic moral worth. Thus, it would appear to be necessary to pin down a single probability measure, or at least a family of similar probability measures, in a non-arbitrary manner.

And here is where things get tricky. What probability measure is the right one for measuring the complexity of *arbitrary* systems? Whatever it is, it must be a probability measure that is in some sense picked out by nature, rather than by our own human interests and concerns. Otherwise, complexity, and thus inherent moral worth, is not really objective, but is tied to a specifically human viewpoint. This goes against the whole thrust of Information Ethics, which seeks to liberate ethics from an anthropocentric viewpoint. Thus, we

need to find a *natural* probability measure for our task. What might such a probability measure look like?

The best-known conception of objective probability is the frequentist conception. According to that conception, the probability of an outcome *O* of an experiment *E* is the proportion of times that *O* occurs in an ideal run of trials of *E*. To apply this notion, we need a well-defined outcome-type *O*, a well-defined experiment-type *E*, and a well-defined set of ideal trials of *E*—and if the latter set is continuous, a well-defined measure on that set. This is all notoriously difficult to apply to non-repeatable event tokens and to particulars in general. To assign a frequentist probability to a particular *x*, it is necessary to subsume *x* under some general type *T*, and different choices of *T* may yield different probabilities. In other words, the frequentist probability of a particular depends among other things on how that particular is described. Different ways of describing a particular will correspond to different conceptions of what it is to repeat that particular, and thus, to different measures of how frequently it occurs in a run of cases.

What this means for us is that the information content of a concrete particular depends, potentially, on how we choose to carve up the world. Again, this is not a problem in practice for information theory, since in any given application, a particular (frequentist) probability measure is likely to be singled out by the problem's context. But in describing the information context of completely arbitrary objects, there is no context to guide us. In particular, if we subsume a concrete particular *x* under a commonly occurring type *T*, it receives a high frequentist probability, and correspondingly low Shannon information content. If we subsume that same particular under a rarely occurring type *T**, it receives a low probability and correspondingly high information content.

Thus, it is by no means obvious that there is a choice of probability measure that (a) is natural independently of our own anthropocentric interests and concerns, and (b) gives us a measure of complexity that is a plausible candidate for inherent moral worth, even assuming that the latter has any special tie to complexity in the first place. To be fair, it is also not obvious that there is not such a probability measure. As the measure *p* from thermodynamics shows, there is at least one natural way of assigning probabilities to physical states, one which does indeed yield a measure of complexity, albeit not the measure of complexity we are looking for. It also raises a further worry. The reason thermodynamic entropy is a bad candidate for basic moral disvalue is simply that it is always increasing, regardless of our actions. That is simply the second law of thermodynamics. What guarantee do we have that complexity, measured in any other way, is not also decreasing inexorably? Thermodynamic entropy can decrease *locally*, in the region of the universe we care about, at the expense of increased entropy somewhere else, and the same may be true for other measures of complexity. But this fact is surely irrelevant to a patient-centered, non-anthropocentric moral theory.

4. Information Everywhere

Statistical information theory is, of course, not the only way to capture the idea of complexity and structure. However, I would argue that the whole notion of complexity or information content becomes trivial unless it is tied to our interests (or someone's interests) as producers and consumers of information.

How much information is there in a glass of water? The obvious, intuitive answer is: very little. A glass of water is fairly homogeneous and uninteresting. Yet the exact state of a glass of water would represent an enormous amount of information if it were described in its entirety. There are approximately 7.5×10^{24} molecules in an eight ounce glass of water.[5] If each molecule has a distinguishable pair of states, call them *A* and *B*, then a glass of water may be regarded as storing over seven trillion terabits of data. Further, let *f* be any function from the water molecules into the set $\{A, B\}$. Relative to *f*, we may regard a given molecule *M* as representing the binary digit 0 if *M* is in state *f*(*M*), and 1 otherwise. Clearly, there is nothing to prevent us from regarding a glass of water in this way if we so choose, and with any encoding function *f* we like. And clearly, by a suitable choice of *f*, we may regard the water as encoding any data we like, up to about seven trillion terabits. For example, by choosing the right encoding function, we may regard the water as storing the entire holdings of the Library of Congress, with plenty of room to spare. Alternatively, a more "natural" coding function, say $f(M) = A$ for all *M*, might be used, resulting in a relatively uninteresting but still vast body of information.

Now if ordinary objects like glasses of water really do contain this much information, then there is too much information in the world for information content to be a useful measure of moral worth. The information we take a special interest in—the structures that are realized in ways that we pay attention to, the information that is stored in ways that we can readily access—is simply swamped by all the information there is. The moral patients we normally take an interest in are vastly outnumbered by the moral patients we routinely ignore. Floridi's estimate of the world's information, a relatively small number of exabytes, is several orders of magnitude lower than the *yottabyte* of information that can be found in a glass of water. Thus, if information content is to serve as a measure of moral worth, the information described in the previous paragraph must be excluded.

But on what basis could it be excluded? We might try to exclude some of the more unconventional encoding functions, such as the encoding function that represents the water as storing the entire Library of Congress. Such encoding functions, it may be argued, are rather unnatural and do not represent the information that is objectively present in the water. Even if this is so, there is no getting around the fact that a glass of water represents a vast amount of information, in that it would take much information to accurately describe its complete state. That information might be rather uninteresting—uninteresting *to us*, that is—but so what? If moral worth is tied to information content per se, then it does not matter whether that information is interesting. If moral worth is tied to *interesting* information, then it appears that moral worth is directly tied to human concerns after all.

But there is a more fundamental problem with dismissing some encoding functions *f* as unnatural. Whenever information is stored in a physical medium, there needs to be an encoding function to relate the medium's physical properties to its informational properties. Often, this function is "natural" in that it relates a natural feature of information (e.g., the value of a binary variable) to a natural feature of the physical medium (e.g., high or low voltage in a circuit, the size and shape of a pit on an optical disk, magnetic field orientation on a magnetic disk, etc.). However, there is absolutely no requirement to use natural encoding functions. There need be no simple relation whatsoever between, say, a file's contents and the physical properties of the media that store the file. The file could be encrypted, fragmented, stored on multiple disks in a RAID, broken up into network packets, etc.

In practice, we always disregard the information that is present, or may be regarded as present via encoding functions, in a glass of water. But the reason does not seem to be a lack of a natural relation between the information and the state of the water. The reason is that even though the information is in some sense there, we cannot easily use or access it. We can regard a

glass of water as storing a Library of Congress, but in practice there is no good reason to do so. By contrast, a file stored in a possibly very complicated way is nonetheless accessible and potentially useful to us.

If this is right, then there is a problem with viewing information's intrinsic value as something independent of our own interests as producers and consumers of information. The problem is that information does not *exist* independently of our (or someone's) interests as producers and consumers of information. Or, alternatively, information exists in an essentially unlimited number of different ways: what we count as information is only a minute subset of all the information there is. Which of these two cases obtains is largely a matter of viewpoint. On the former view, even if inanimate information has moral value, it has value in a way that is more tied to a human perspective than Floridi lets on. On the latter, there is simply too much information in the world for our actions to have any net effect on it.

5. Conclusion

The immediate lesson of the last two sections is that overall complexity, or quantity of information, is a poor measure of intrinsic moral worth. Now this conclusion, even if true, may not appear to be terribly damaging to Information Ethics, as the latter embodies no specific theory of how to measure moral worth. It may simply be that some other measure is called for. However, I would argue that the above considerations pose a challenge to any version of Information Ethics, for the following reason.

As we have seen, the number of (informational) objects with which we interact routinely is essentially unlimited, or at least unimaginably vast. If each object has its own inherent moral worth, what prevents the huge number of informational objects that we do not care about from outweighing the relatively small number that we do care about, in any given moral decision? For example, I might radically alter the information content of a glass of water by drinking it, affecting ever so many informational objects; why does that fact carry less moral weight than the fact that drinking the water will quench my thirst and hydrate me? The answer must be that virtually all informational objects have negligible moral value, and, indeed, Floridi seems to acknowledge this by saying that many informational objects have "minimal" and "overridable" value. But that claim is rather empty unless some basis is provided for distinguishing the few objects with much value from the many with little value.

Of course, one answer is simply to assign moral worth to objects based on *how much we care about them*. That would just about solve the problem. Moreover, that is more or less what it would *take* to solve the problem, insofar as the objects that must be assigned minimal value (lest ethics become trivial) are in fact objects that we do not care about. However, this is not an answer Floridi can give. Moral worth is supposed to be something objects possess intrinsically, as parts of nature. It is not supposed to be dependent on our interests and concerns. Thus, what is needed is an independent standard of moral worth for arbitrary objects which, while not based directly on human concern, is at least roughly in line with human concern. And so far that has not been done.

Endnotes

1. See, for example, Floridi 2007, Floridi 2008a, etc.
2. See Floridi 2008b.
3. See Shannon 1948. For a good modern introduction, see MacKay 2003.
4. A base-2 logarithm is used because information is measured in bits, or base-2 digits. If information is to be measured in base-10 (decimal) digits, then a base-10 logarithm should be used. In general, the Shannon information content is defined to be $\log_b(1/p(M))$, with b determined by the units in which information is measured (bits, decimal digits, etc.).
5. This figure is obtained from the number of molecules in a mole (viz. Avogadro's number, approximately 6×10^{23}), the number of grams in one mole of water (equal to water's atomic weight, approximately 18), and the number of grams in 8 ounces (about 227).

References

Floridi, Luciano. 2008a. Information Ethics, its Nature and Scope. In *Moral Philosophy and Information Technology*, edited by Jeroen van den Hoven and John Weckert. Cambridge: Cambridge University Press.

———. 2008b. A Defence of Informational Structural Realism. *Synthese* 161(2):219-53.

———. 2007. Understanding Information Ethics. *APA Newsletter On Philosophy and Computers* 07(1):3-12.

MacKay, David J.C. 2003. *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.

Shannon, Claude. 1948. A Mathematical Theory of Computation. *Bell System Technical Journal* 27:379-423 and 623-56.

Understanding Luciano Floridi's Metaphysical Theory of Information Ethics: A Critical Appraisal and an Alternative Neo-Gewirthian Information Ethics

Edward Howlett Spence
University of Twente, Netherlands

1. Floridi's Information Ethics

Being beyond the scope of this short paper and unavoidably constrained by space, I can but offer the briefest of expositions of Floridi's rich and complex theory, but hopefully I can at least provide in a summarized form the direction and main rationale of that theory and importantly not misconstrue it in the process. In addition, I shall offer some well intentioned and hopefully helpful critical observations and then proceed to offer an alternative approach to IE based on Alan Gewirth's rationalist ethical theory, specifically his argument for the foundational moral principle of morality, the Principle of Generic Consistency (PGC), extended and adapted for that purpose.

Beginning with the uncontroversial empirical observation that our society is evolving, both quantitatively and qualitatively, into an information society, Floridi introduces the concept of *infosphere*, the informational equivalent of "biosphere." According to Floridi, *infosphere*

> Denotes the whole informational environment constituted by all informational entities. …It is an (intended) shift from a semantic (the infosphere understood as a space of contents) to an ontic conception (the infosphere understood as an environment populated by informational entities)." (Floridi 2007, 4)

Floridi goes on to claim that this informational shift from the semantic to the ontic is resulting in the *re-ontologization* of the world that "transforms its intrinsic nature" (Floridi 2007, 4) so that the world can now be ontologically re-conceived according to Floridi as being fundamentally constituted by the infosphere and not merely the biosphere, as was previously thought. As an example, he cites nanotechnologies and biotechnologies that "are not merely changing (re-engineering) the world in a very significant way (as did the invention of gunpowder, for example, but actually reshaping (re-ontologizing) it" (Floridi 2007, 4).

As a result of this ontologization, information is becoming our ecosystem and we, together and in interaction with artificial agents, are evolving into informationally integrated *inforgs* or *connected informational organisms* (Floridi 2007, 5-6). Floridi predicts that "in such an environment, the moral status and accountability of artificial agents will become an ever more challenging issue" (Floridi 2007, 5).

From this initial ontological thesis, namely, the ontologization of the infosphere or the metaphysics of information, it is easy to anticipate Floridi's next theoretical move. On the basis of his metaphysics of information Floridi posits a "new environmental ethics" when information ethics ceases to be merely "*microethics* (a practical, field-dependent, applied, and professional ethics)" and becomes instead "a *patient-orientated*, *ontocentric* {as opposed to merely biocentric}, *ecological* macroethics" (Floridi 2007, 7-8). "Information ethics is an ecological ethics that replaces biocentrism with ontocentrism," a substitution in the concept of biocentrism of the term "life" with that of "existence" (Floridi 2007, 8). This substitution, as we shall see below, is both crucial and problematic in Floridi's overall thesis of Information Ethics.

The claim that information ethics can be conceived and ought to be conceived as an environmental macroethics is Floridi's most interesting, ambitious, and challenging claim in his theory and constitutes the crux of his whole controversial argument that rightly or wrongly is conducive to raising many incredulous stares. For the claim amounts to nothing less than the clear implication, as expressed openly by Floridi himself, that existence, not life, is the mark of morality; that which determines the moral status of not only humans and other sentient beings, including their natural environment—the whole biosphere, but moreover, at the most ultimate level of inclusiveness ever conceived in moral philosophy before, the moral status of the whole caboodle, everything that exists, has existed, and ever will exist in the Universe as informational objects. Which, essentially, insofar as anything can be conceived as an informational object, means practically *everything*, including artefacts, works of art, gardening tools, coffee mugs, tea-cups, carpets, pebbles, rocks, clarinets, and, if I am not mistaken, kitchen utensils such as knives, for example. This is an *ethics of being* on a grand scale that considers the *destruction, corruption, pollution, and depletion* of informational objects as a form of *entropy* whose increase constitutes an instance of evil that should, all things being equal, be ethically avoided (Floridi 2007, 9).

> In IE, the Ethical discourse concerns any entity, understood informationally, that is, not only all persons, their cultivation, well being, and social interactions, not only animals, plants, and their proper natural life, but also anything that exists, from paintings and books to stars and stones; anything that may or will exist, like future generations; and anything that was but is no more, like our ancestors or old civilizations. Information Ethics is impartial and universal because it brings to ultimate completion the process of enlargement of the concept of what may count as a centre of a (no matter how minimal) moral claim, which now includes every instance of *being* understood informationally, no matter whether physically implemented or not. In this respect, IE holds that every entity, as an expression of *being*, has a dignity, constituted by its mode of existence and essence… (Floridi 2007, 9)

The above evocative passage encapsulates the essential characteristics of Floridi's Information Ethics and illustrates its extensive scope. It is, as Floridi states, a universal ethics that applies equally to all informational objects in the Universe. I will go as far as saying that it seems to offer a kind of Stoic Pantheistic Ethics (my phrase) that endows everything in the Universe with a moral significance and status through a pre-determined divine rational order in which everything is ontologically inter-connected and of which everything forms an ontic part, no matter how big or small.

2. Some Sceptical Observations

It seems that, according to Floridi, the basis of having a moral status is the *informational state* possessed by an entity (Floridi 2007, 10). Insofar as all entities whether sentient or non-sentient can be conceived as having this informational state, then they are entitled to a moral status:

> The result is that all entities, qua informational objects, have an intrinsic moral value, although possibly quite minimal and overridable, and hence they can count as moral patients, subject to some equally minimal degree of moral respect understood as a *disinterested, appreciative, and careful attention*…There seems to be no good reason not to adopt a higher and more inclusive, ontocentric perspective. (Floridi 2007, 10)

I agree with Floridi that there would be no good reason not to adopt such a higher and more inclusive moral perspective if there were in fact good objective and independently grounded reasons for adopting such a perspective. This would in fact be a welcome extension to the moral fabric of the world. But merely declaring such a moral status for all informational objects on the basis of their informational state alone does not constitute such justified reasons. That is to say, the informational status of the informational objects cannot of itself provide them with a moral status any more than the human status of people can of itself provide them with a moral status.

By contrast, Alan Gewirth's Principle of Generic Consistency (PGC)[1] that was briefly cited above could be applied to argue that the natural property of purposive agency that acts as the sufficient condition for having rights to freedom and well being can be extended to purposive agents and patients other than human beings, for example, to animals and androids. Insofar as animals and other sentient beings can be said to posses some degree of purposive and goal orientated behavior that requires them to posses some minimal degree of freedom and well being, they too are entitled to rights to freedom and well being as patients if not as agents. For insofar as one recognizes that animals and other sentient beings possess purposive agency, minimal as that may be, and that this alone is a sufficient condition for granting them a moral status, one must at least rationally acknowledge that they too have rights to freedom and well being, at least as patients, on pain of self-contradiction. Some similar argument is also required for extending the moral status to non-sentient informational objects and inforgs. But what could it be?

My reading of Floridi suggests that ontic existence alone qua informational object suffices to establish the moral status of the informational object. But why is this so? How can existence of itself entitle an entity including human beings to a moral status? What is required in establishing such a claim is to show that ontic existence per se, and in particular ontic existence qua informational object, endows one with intrinsic value and thus a moral status. But how can a morally neutral and value-neutral ontological property such as existence confer of itself moral value and moral status to the entity that possesses it, be it sentient or non-sentient?

Unless justified reasons can be provided that lend support to the claims (a) that the mere existence of non-sentient entities as informational objects renders them intrinsically valuable

by virtue of existence itself having an intrinsic value, which by extension is bestowed on anything that possess it, be it sentient or non-sentient; and (b) by virtue of the possession of that intrinsic value, informational objects of any kind should be accorded a minimal moral status; the claim that non-sentient informational objects have an intrinsic value and hence a minimal moral status cannot be sustained. If I am not mistaken, I do not think Floridi has provided independently justified reasons in support of the two claims referred to above, namely, that existence per se has an intrinsic normative value, which by extension is bestowed on anything and everything that possess it as informational objects or inforgs, be they sentient or non-sentient.

Floridi seems to merely assert that existence has an intrinsic value at the threshold of some Level of Abstraction (LoA) which confers a moral status on all that possess it (all informational entities both sentient and non-sentient) without providing independent and objective justified reasons to demonstrate why this is the case. He does, however, state that "moral agenthood," and I suppose by parity of argument moral patienthood as well, "depends on a LoA." He goes on to say that "morality may be thought of as a 'threshold' defined on the observables in the interface determining the LoA under consideration" (Floridi and Sanders 2004, *On the Morality of Artificial Agents*). The question, however, arises of how and why on the basis of such a LoA at some defined observable threshold does the conferring of intrinsic moral value to all informational objects take place and does so independently of any anthropomorphic perspective.

Unless I have misunderstood how the LoA is applied in the attribution of moral value to informational objects and moral agency in general, the conferring of moral value to informational objects, especially of the non-sentient kind, seems rather mysterious. Isn't the LoA part of the anthropomorphic perspective? If it is, how is the moral value attributable to informational objects independent of that perspective, which it needs to be, if moral value is to be attributed to them from outside an anthropocentric perspective that only intrinsically and unconditionally valuable entities such as sentient beings and their supporting environments can be said to have?

Moreover, how can the choice of a LoA, which of itself is a value-neutral concept, generate the moral value of informational objects? If it is the *choice* of the LoA that is the basis for conferring moral value to informational objects and not the LoA itself, how does that *choice* of itself confer value to informational objects and importantly, what are the reasons for thinking that the *choice* of a LoA is able of itself to confer moral status to informational objects? What needs to be shown is that the LoA is somehow a morally conferring property or concept, independently of the choice made of that LoA or that the choice itself of that LoA somehow of itself confers a minimal moral value to informational objects. However, I do not think that this has been shown, at least not by justified reasons based on some independently objective argument.

3. Information Ethics without Metaphysics

By contrast to existence, purposive or goal-orientated behavior can confer value in the manner demonstrated by Alan Gewirth's argument for the PGC (1978). Namely, the necessary conditions for purposive agency, freedom, and well being, which are also necessary for a meaningful and worthwhile life, provide the basis for having rights to freedom and well being and hence provide the universal foundation for the moral status of all purposive agents or patients, be they human or non-human.

One way to extend the moral status to non-sentient informational objects could be accomplished by showing how non-sentient informational objects possess in some sense and to some degree a form of purposive agency or some other teleological property that is value conferring. Insofar as information can be said to be goal-orientated or teleological in some relevant sense, this might not prove impossible, difficult though as it might seem at present.

Consider this argument. I will refer to it as the *Argument from Designed–in-Purposive Agency* (*A-DiPA*). Artefacts and other non-sentient informational objects have a functional instrumentality. They are designed to perform a certain specific functional and instrumental role. Take a knife, for example. The functional role of a knife is to cut materials of a certain kind. It has been designed with that functional purpose in mind. This functional role or purpose is inherently designed in the knife and as such *inheres* in the knife unless removed. All things being equal the knife when used as intended will cut perfectly well according to the purpose for which it was designed—its design-in-purpose. Now let us suppose that someone *for no good reason* and merely on a whim destroys the teleological (its design-in-purpose) and functional capacity of the knife to cut. Let us also assume that this someone, call him Mack, is the owner of the knife. The knife is now blunt and has lost its functional purpose of cutting. No doubt the knife has been *damaged* (harmed) instrumentally as it can no longer fulfill the instrumental role or the purpose for which it was designed and created. But has any moral harm been committed and, if so, to whom and by whom?

To answer this question let us first ask a different question: Would it have been better if Mack had not and for no good reason destroyed the capacity of a perfectly good knife to cut? If the answer to that question is yes, as it is likely to be, we can then proceed and ask what kind of damage or harm has been committed. I think we can allow that an instrumental harm has taken place which would have been better had it not occurred. What about a moral harm? Has the knife suffered a moral harm by it being made blunt? Clearly not as an agent, since the knife lacks the capacity for agency. Following Floridi and Sanders (2004, 349) the knife can be said to lack agency because it lacks its three essential features of interactivity (*response to stimulus by change of state*), autonomy (*ability to change state without stimulus*), and adaptability (*ability to change the 'transitions rules' by which state is changed*).

However, even if the knife lacks the capacity for agency in the full-blooded and traditional sense, could we not argue that the knife because of its inherent or *designed-in-purposiveness* or *designed-in-teleology* has some other type of distributed agency (Floridi and Sanders 2004, 351) or *contributive* agency (Korsgaard 1983, 172), which affords it some minimal moral role? After all, a knife can be used to murder, a typical immoral action. Let us assume that if the murderer had not possessed a knife they would not have been able to commit the murder, and thus an immoral act would not have taken place. Under this assumption, the knife can be said to have contributed to the murder in virtue of its inherent teleology or *designed–in-purposive-agency* (DiPA), or that the immoral act of the murder can be defined as *morally distributed* across a *moral-field* or *moral-network* that at least includes the murderer (the prime moral agent), the teleological instrument (the knife as a morally contributing *and* instrumental agent), and the victim (the moral patient). Following Floridi and Sanders (2004, 366-69), I will argue that although the knife can of course not be held in any way morally *responsible* for the murder it can nevertheless be held *accountable* in virtue of its contributed role to the murder via its designed-in-purposive-agency or DiPA. There is, as Floridi and Sanders rightly claim, a conceptual difference between moral responsibility and moral accountability. Although an earthquake can be held accountable for the moral harm of its

victims as the primary *cause* of that harm it cannot, because it lacks the relevant full-blooded agency, be held morally responsible.

Adapting and extending Gewirth's argument from the Principle of Generic Consistency on the basis of which it is shown that purposive agents have rights to freedom and well being for the sufficient reason that they are purposive agents (that is, they possess the natural property of purposive agency), can we not *reasonably* say that artefacts such as knives with a designed-in-purposive-agency (the designed-in goal or purpose to do x, in the case of the knife, x = to cut) have to some minimal degree *prima facie* rights to (Art)freedom (artificial freedom) and (Art)well being (artificial well being) as patients if not as agents? That is to say, can we not reasonably say that such artefacts have the right not to have their (Art)freedom in exercising their designed-in-purposive-agency thwarted or interfered with for no good reason, or their (Art)well being violated by having their DiPA, within which their (Art)well being can be defined and understood in terms of *what they are good for* (their designed-in "functional goodness" or "designed-in-capacity" to do x) reduced or eliminated for no good reason?

Can we not say, following this line of thinking, that Mack's knife that was rendered useless by being made blunt for no good reason had its (Art)freedom and (Art)well being unjustifiably violated and thus suffered not only an instrumental harm by having its instrumental functional role damaged, but also a moral harm qua artefact worthy of some minimal respect owed to it by virtue of its DiPA? Although the instrumental role of the knife can be replaced by the replacement of the damaged knife by a new one, the knife itself that was made blunt for no good reason has not only lost its replaceable instrumental functionality but also its irreplaceable particular inherent capacity to do what it was designed to do best, namely, cut well. That inherent capacity is something that the knife possessed as a thing-in-itself and as such it is something that can be valued for its own sake and not merely instrumentally for the sake of being able to cut well for some human agent.

Following Korsgaard's distinction between objective intrinsic and unconditional value on the one hand and objective but extrinsic conditional value on the other (1983), I will argue that the knife has suffered moral harm by being damaged; that is, by having its DiPA to cut well rendered useless.

According to Korsgaard something X has an objective extrinsic but conditional value if X meets the relevant conditions under which it is held to be valuable and X is also something that is valued for its own sake or as an end, and *in addition* to its instrumentality as a means (1983, 184ff). Going along with Korsgaard we can then say that a knife or other relevant informational object is valued or can be valued partly for its own sake as an end in addition to its instrumental use as a means for human ends, provided certain relevant conditions are met. For example, that when a knife is used it is used for good ends and not for bad ends. Having this dual value, both instrumental as a means and extrinsic or inherent value[2] as an end, the instrumental disvalue of a knife or other object that is being used to commit a moral wrong diminishes and tramps its inherent value as an end. This follows from the fact that the knife and other objects of this ontological type only have conditional value so that it would be justified to destroy a perfectly good knife if that were the only way to prevent a murder, for example.

In the case of Mack's knife, by contrast, both the extrinsic and instrumental value of the knife have been diminished, eliminated, in fact, *for no good reason*; that is to say the conditions under which the knife is considered or can be considered valuable have been violated by the blunting of the knife, *for no good reason*. The qualification *for no good reason* is crucial and seems to point in the opposite direction in which Floridi's argument for assigning moral value to informational objects seems to go. For I am partly in agreement with Korsgaard although for Gewirthian reasons rather than Kantian as in her case, that the objective and inherent value or for Korsgaard extrinsic value of an object, or informational object as in Floridi's case, is not just a matter of the ontological status of the object qua informational object but of practical reason as well (Korsgaard 1983, 183-84). I said I am only *partly* in agreement with Korsgaard because her claim is that the extrinsic value or, in my case, inherent value of an object is only a matter of practical reason and not one of ontology. Orientating my own position somewhere between that of Korsgaard and Floridi, I want to argue that the value of an object and in particular an informational object is determined partly by its ontology by virtue of its designed-in-purposive-agency (DiPA)—the artificial equivalent of the natural property of purposive agency inherent in human beings and some other animals—and partly by the reasons we have for holding that artefact valuable, principally, in virtue of the reasons for which we hold artefacts of a certain kind to be good for doing x, by virtue of possessing the capacity to fulfill certain designed-in goals or purposes for doing x.

That is to say, what drives us to attribute objective but conditional value to an informational object as a thing valued for its own sake and not merely as an instrument for advancing our own ends, such as a knife, for example, are partly the reasons themselves for designing such objects. The value or goodness of those reasons is transferred through the designing and creation of those objects into the objects themselves. Through this transference of *reasonable* value into the objects on the basis of the functional excellence and efficacy of their designed-in-agency or functional teleology, the value transferred through the design of the objects persists to inhere in the objects until the conditions under which those reasons hold valuable and good are diminished or eliminated as when a knife designed for cutting bread is used to commit murder, for example. Note that a gun used to kill in self-defense does not have its inherent value diminished by its instrumental use where by contrast a gun used to murder does; that is, the instrumental disvalue of murdering someone diminishes or eliminates the inherent objective value of the gun.

Insofar as a knife can be said to have an inherent value or what Korsgaard defines as an objective extrinsic but conditional value, and insofar as Mack knife's value has been eliminated for no good reason (the relevant condition in this case), the elimination or diminution of the value of the knife or of any other teleological object can be said to be a moral harm. For the unreasonable elimination or diminution of an objective inherent or objective extrinsic conditional value is unjustified (because no good objective reason can be given for it) and hence morally wrong as it diminishes value overall. In the case of Mack's knife it diminished both instrumental and inherent value as the knife in its prime condition possesses both. It has the instrumental value of being used as a perfectly good knife to cut, an apple, for example, but it also possesses an inherent designed-in-purposive-capacity to cut whether or not it is ever used in that way. A good knife that lay dormant and was not used to cut would retain that inherent value regardless of whether their designed-in-purposive-capacity was put to instrumental use or not. And it is this conceptual distinction just made between the knife's *in-use-instrumental-value* exercised in cutting things and its inherent value, which it has by virtue of its *designed-in-purposive-value* that affords it the capacity to cut, that allows us to ascribe to the knife and other objects or artefacts of the type that possess a designed-in-purposive-agency (DiPA), two inter-related values: one instrumental and one inherent.

4. Implications for Floridi's Ontological Thesis for the Moral Value of Informational Objects

In his paper "On the intrinsic value of information objects and the infosphere" (2002) Floridi postulates the two theses that comprise his Information Ethics (IE) theory:

1. *The first thesis states that information objects qua information objects can be moral agents.*
2. *The second thesis states that information objects qua information objects can have an intrinsic moral value, although quite minimal, and hence that they can be moral patients, subject to some equally minimal degree of moral respect.*

My analysis above in terms of attributing inherent but conditional moral value to informational objects such as a knife, for example, seems to support both of Floridi's two theses of IE but without the metaphysical cost of having to postulate two extra metaphysical claims to the effect that (a) anything that exists in the infosphere as an informational object has moral value just by virtue of its ontic existence and (b) the unjustified damage or destruction of informational objects due to a lack of respect for their minimal moral worth causes information entropy, which is overall a bad outcome and one that ought to be avoided.

I have argued above that existence per se even qua information objects cannot of itself confer moral value. Floridi's motivation for choosing the primary ontological route to the moral worth of informational objects is that he thinks that existing ethical theories which are either predominantly anthropocentric such as Kant's theory, or various other biocentric theories which are more inclusive than Kant's theory but not sufficiently so, cannot account for the moral worth of non-sentient objects such as artificial systems like software agents in cyberspace (2002, 299), for example. If my analysis above is correct, Floridi's motivation is justified but misdirected. Justified because he is right in arguing that there is a theoretical need to extend the moral sphere to include not just all sentient and other living organisms in the biosphere but also all entities that qualify as information objects including non-sentient beings such as coffee mugs, knives, and software agents or webbots (Floridi and Sanders 2004, 370) in the infosphere. As he states, "*showing that both an anthropocentric and biocentric axiology are unsatisfactory is a crucial step*" (2002, 291).

However justified his motivation for extending the moral sphere to include not only the biosphere but also the infosphere is, the exclusive *ontocentric* orientation of his approach in seeking to confer moral value to information objects merely on the basis of their existence is misdirected because it lacks sufficient justification and the justification if any that it does have comes at a higher metaphysical cost than what is required. Ockham's razor counsels against ontological inflation and for metaphysical economy.

My Neo-Gewirthian approach, which locates the inherent moral worth and value of all informational objects, including human beings, animals, and inanimate objects such as artefacts, the whole of Floridi's infosphere in fact, in the natural property of *purposive agency* provides, I believe, adequate justification at no additional ontological cost. Contrary to Floridi whose profound insights into the meta-theoretical need for attributing moral value to all informational objects qua informational objects I share, I have argued that we do not require additional ontological categories or extra metaphysical machinery for doing so. The capacity for purposive agency alone, which is the natural property on the basis of which human beings and other sentient beings such as animals have inherent moral worth, can be adapted and extended, as I have shown above, to include other non-sentient information objects, such as knives, for example. Whereas sentient beings posses purposive agency naturally and inherently by varying degrees from very high in the case of human beings and perhaps high in the case of dolphins and whales to very low in the case of amoebas, non-sentient beings such as artificial agents on the higher scale and thermostats and knives on a lower scale possess an *artificial purposive agency* by design and teleological implantation that *inheres* in those objects and renders them inherently but conditionally morally valuable as I have argued above. By extension of Gewirth's argument for the Principle of Generic Consistency, they have rights to (Art)freedom (artificial freedom) and (Art)well being (artificial well being).

My Neo-Gewirthian approach of attributing inherent moral worth to all informational objects as entities to be valued for their own sake as ends in themselves, unconditionally with regard to human beings but conditionally with regard to non-sentient entities such as knives and other teleological artefacts, is partly in agreement with Floridi's claims that

> *There seems to be no good reason not to adopt a higher and more inclusive, ontocentric LoA*. (2002, 291)

and that

> *The moral worth of an entity is based in its ontology. What the entity is determines the degree of moral value it enjoys, if any, whether and how it deserves to be respected and hence what kind of moral claims it can have on the agent*. (2002, 294)

I say only partly because although purposive agency as the basis of all moral worth is itself an ontological category, it has the advantage of comprising a natural property with no need to introduce additional and costly metaphysical theoretical postulates to explain the moral worth of informational objects as Floridi does. For the capacity for purposive agency as the basis for attributing moral worth to an entity qua informational object, to some varying degree, is sufficient in explaining and accounting for the moral worth of both sentient beings, organisms, and systems that inherently posses the capacity for purposive agency naturally, and non-sentient entities such as artificial agents, for example, that possess the capacity for purposive agency contributively through having it artificially designed and implanted in them, by human agency. However, once implanted, that capacity for contributive purposive agency, which I named earlier in the paper as Designed-in-Purposive-Agency (DiPA), becomes and remains inherent within the non-sentient entity until removed or eliminated, again by human design.

As an inherent property, it has a moral value, both instrumentally and inherently as explained above (it has a dual value) that is independent of the wishes or sentiments of any particular human agent. A good knife is a good knife (one that has the capacity to cut well as designed to do) whether one wishes it or not, or whether or not it serves any particular human interest. Of course, if the use of knives for cutting became completely redundant and obsolete, they would lose the inherent and instrumental value that they now possess. It is for that reason that in agreement with Korsgaard I also wish to claim that the value of artificial entities such as knives and coffee mugs is partly conditional on their factional usefulness and their perceived value based on practical reason for which they were designed. But in disagreement with Korsgaard's Kantian perspective I wish to claim that this is, however, different in the case of sentient beings, such as animals, for example, that retain their inherent moral value regardless of whether or not they have any functional use or value for human beings. Cows

that can no longer produce milk or chickens that no longer lay eggs are still morally worthy of consideration in their own right regardless of human needs and interests.

This final point seems to accord with Floridi's own claim that

> *It seems reasonable to assume that different entities may have different degrees of relative value that can constrain a's [the agent's] behaviour without necessarily having an instrumental value, i.e., a value relative to human feelings, impulses or inclinations, as Kant would phrase it*. (2002, 293)

Although the capacity for purposive agency both naturally in the case of sentient entities and artificially in the case of non-sentient entities creates a continuum of moral worthiness and moral consideration across a wide network of informational objects, that continuum is separated by qualitative divisions between those entities that affords them various differentiated degrees of moral value in terms of the complexity of their capacity for purposive agency. Using the metaphor of canal or river locks we can say that because the moral continuum of informational objects is porous, the capacity of purposive agency slips through the various qualitative moral divisions like water through the locks in a canal or river. However, the transitions from one qualitative moral division to another requires, as in the case of the raising of the water level in a lock to allow a ship to transit from one level of the canal to another, the raising of the level of complexity of an entity's capacity for purposive agency so as to enable its transition from a lower to a higher qualitative moral division. Thus, a software agent's capacity for purposive agency would have to be raised to that of an intelligent android that meets Floridi's and Sanders' conditions of full agency discussed above before it can proceed to a higher moral division close to that of human beings.

The conceptual distinctions between on the one hand responsibility and agenthood and on the other accountability and patienthood help explain the relative moral value of different entities. Thus, although we could only hold a software agent accountable but not responsible for the destruction of valuable information, we could by contrast hold an android or human agent both accountable and responsible due largely to their higher moral status. Similarly, although we ought to morally avoid killing a tiger unless in self-defense we cannot reasonably expect a tiger to morally reciprocate in the same moral way. This is because although a moral patient worthy of moral respect the tiger does not posses sufficient moral agency to warrant us holding the tiger bound to reciprocal moral obligations with regard to human agents. Thus, the four conceptual distinctions of responsibility/accountability and agenthood/patienthood go some way in explaining the relative moral value of different informational objects in relation to the moral relevance and significance of those conceptual categories in specific contexts.

6. Conclusion

Floridi's reference to Spinoza (2007, 9) seems to suggest that he may be entertaining, not explicitly but perhaps implicitly, a Stoic perspective with regard to his metaphysical thesis of Information Ethics. My alternative Neo-Gewirthian thesis of Information Ethics gives a more explicit expression to that implied suggestion. For my thesis is based on the view supported by argument that the moral value of all entities at least on Earth, both sentient and non-sentient, is comprised of a composite dual nature or double-aspect nature of being at once *purposive-agentive entities* as well as entities imbued with *reason*: in the case of sentient beings, intrinsically and unconditionally, and in the case of non-sentient beings (such as artefacts) inherently but conditionally by *rational design*. It is my claim that it is this composite dual character that allows for the attribution of moral value to all entities, both sentient and non-sentient. This analysis is in keeping with a claim I make elsewhere that Gewirth's rationalist ethics and in particular my Neo-Gewirthian reconstruction and expansion of it, is essentially Neo-Stoic.[3]

Finally, there might be other necessary reasons of why Floridi introduces the machinery of his ontological metaphysics (see, for example, his "Informational Structural Realism," 2008, *Synthese*) but this cannot be for establishing the moral worth of informational objects because, as I hope to have demonstrated, none is required.

Endnotes

1. Due to constraints of space, I will not be able to provide a justification for Alan Gewirth's argument for the Principle of generic Consistency (PGC) on which his derivation of rights to freedom and well being is based, as this is well beyond the scope and limits of this paper. I offer such a detailed defense in my *Ethics Within Reason: A Neo-Gewirthian Approach* (2006).
2. I prefer to use the term *inherent* rather than Korsgaard's *extrinsic* term because the value an artefact has by virtue of its DiPA inheres in the artefact and so it is not exclusively determined by the external reasons for which human beings hold it to be valuable. I should add, however, and perhaps this is in keeping with Korsgaard's position, that in the event that an artefact was no longer held to be valuable its inherent value by virtue of its DiPA could be revoked. For what can be designed in can also be designed-out. This is in keeping with the correct thought that values are to a large degree determined by the underlying reasons for considering those values "valuable."
3. See Chapter 10 of Spence 2006, *Ethics Within Reason: A Neo-Gewithian Approach*, 393-442.

References

Floridi, Luciano. 2008. Informational Structural Realism. *Synthese* 161(2):219-53.

———. 2007. Understanding Information Ethics. *APA Newsletter on Philosophy and Computers* 07(1):3-12.

———. 2005. Is Semantic Information Meaningful Data? *Philosophy and Phenomenological Research* LXX(2).

———. 2004. On the Morality of Artificial Agents. Minds and Machine 14:349-79.

———. 2002. What is the Philosophy of Information? *Metaphilosophy* 33:123-45.

———. 2002. On the Intrinsic Value of Information Objects and the Infosphere. *Ethics and Information Technology* 4:287-304.

Gewirth, A. 1978. *Reason and Morality*. Chicago: University of Chicago Press.

———. 1996. *The Community of Rights*. Chicago: University of Chicago Press.

———. 1998. *Self-fulfillment*. NJ: Princeton University Press.

Korsgaard, Christine M. 1983. Two Distinctions in Goodness. The Philosophical Review 92(2) 169-95.

Spence, E. 2007. *What's Right and Good about Internet Information? A Universal Model for Evaluating the Cultural Quality of Digital Information*. In *Proceedings of CEPE 2007, The 7th International Conference of Computer Ethics: Philosophical Enquiry*, edited by Larry Hinman, Philip Brey, Luciano Floridi, Frances Grodzinsky, and Lucas Introna. University of San Diego, USA, July 12-14 2007, ISSN 0929-0672.

———. 2006. *Ethics Within Reason: A Neo-Gewirthian Approach*. Lanham: Lexington Books (a division of Rowman & Littlefield).

Discussion Articles on Baker

Artifacts and Mind-Independence: Comments on Lynne Rudder Baker's "The Shrinking Difference between Artifacts and Natural Objects"

Amie L. Thomasson
University of Miami

Against contemporary reductivist and eliminativist trends, Lynne Baker argues in "The Shrinking Difference between Artifacts and Natural Objects" (2008, following up on her 2007 book) that artifacts should be considered just as genuine parts of our world as natural objects are. I couldn't agree more. Of five different ways in which one might attempt to distinguish artifacts and natural objects, she argues, four fail to distinguish them, and the fifth, while distinguishing them, does not warrant denying that artifacts are "genuine substances" (2008, 3-4).

The fifth criterion, which Baker admits does distinguish artifacts from natural objects, is that the "identity and persistence" of artifacts depends on human intentions (2008, 3). This fifth criterion admits of (at least) two interpretations, given two different senses in which artifacts are apparently mind-dependent. First, individual artifacts are existentially mind-dependent in the sense that no table, painting, or computer could exist in a world absent of human intentions—in Baker's terms, they are "Intention-Dependent" objects, such that, as she puts it, "the existence of artifacts depends on us" (2008, 4). This intention-dependence, moreover, is not just a causal matter but a conceptual matter or metaphysical matter: the very idea of an artifact is the idea of an *intended* product of human intentionality (cf. Thomasson forthcoming). Second, artifactual *kinds* (such as *chair*, *fork*, and *house*) are often thought to be mind-dependent in the sense that what it takes for there to be members of the kind, and under what conditions members of the kind come into existence and cease to exist, are determined by conditions we *accept* as relevant, rather than forming discoverable features of the world (as the parallel conditions for natural kinds are supposed to). As Baker puts it, the "conditions of membership" in the substance-kind are set by us (2008, 4). So let me add to her case by arguing that *neither* sense of dependence should lead us to deny that artifacts are real parts of our world.

The first sense of dependence is that individual artifacts are existentially dependent on human intentions. But there are important differences among existentially mind-dependent entities. Imaginary objects might be said to be existentially mind-dependent (if they are allowed to exist at all), but they are the products *merely* of human thoughts and intentions. By contrast, artifacts such as tables and chairs cannot be brought into existence by thought alone, but also require real physical acts of hammering, assembling, etc., and depend on their material bases as well as on the human intentions that (e.g.) endow them with a function. This alone should help undermine the idea that allowing the existence of any kind of mind-dependent objects involves countenancing "magical modes of creation" (cf. my forthcoming).

Moreover, the thought that *any* mind-dependence undermines an (alleged) entity's claim to existence is based on illegitimately generalizing from the case of scientific entities: If we found out that some posited scientific entity (say, a planet or a species of bird) was really just "made up," a human creation, we might indeed have reason to say that Vulcan (or the Key Sparrow) doesn't exist. But that reflects the fact that planets and animals are supposed to be mind-*independent*. The same does not go for artifacts: the very idea of an artifact (or work of art, fictional character, or belief or desire) is of a human creation, and so the fact that (e.g.) a table could not have existed were it not for the relevant human intentions does nothing to undermine its claim to existence.

Thus, a mind-independence criterion may be suitable for would-be natural objects, but not for artifacts (or many other sorts of thing). Since the very idea of an artifact is of something mind-dependent in certain ways, accepting mind-independence as an across-the-board criterion for existence gives us no *reason* to deny the existence of artifacts; it merely begs the question against them (see my forthcoming). In fact, considering artifacts gives us reason to be suspicious of proposals for across the board criteria for existence and suggests that we should instead address existence questions separately, asking in each case what it would take for there to be objects of the kind and then determining whether or not those conditions are fulfilled—while acknowledging that criteria for existence may vary for different sorts of thing (cf. my forthcoming).

The second, perhaps more controversial, sense of dependence is the sense in which the conditions for membership in an artifactual kind, and for the existence, identity, and persistence of its members, are themselves mind-dependent. For, as I have argued elsewhere (2003; 2007), what distinguishes the natures of artifactual kinds from those of chemical or biological kinds is (roughly) that we—the makers and users of artifacts of various kinds—determine what features are and are not essential to being a member of an artifactual kind (like *chair*, *split-level*, or *convertible*), in a way that we do not determine the particular features relevant to being a member of a natural kind (like *tiger* or *gold*).

It is often held, however, that possessing a nature that is entirely independent of human concepts, language, etc., which is open to genuine discovery and about which everyone may turn out to be ignorant or in error, is a central criterion for treating kinds as real or genuine parts of our world (Elder 1989, Lakoff 1987). If that's right, we're left with the options of giving up an ontology of artifactual kinds or giving up the idea that possessing discoverable mind-independent natures is the central criterion for "really" existing.

I have argued elsewhere (forthcoming) in favor of the latter route. The thought that, to be real, artifactual kinds must have mind-independent *natures* again comes from borrowing an idea suitable for realism about *natural* kinds and assuming it must apply wholesale. For while *natural* kinds may have to have mind-independently discoverable particular natures, to require this of artifactual kinds misconstrues what it is to be a realist about *artifactual* kinds. For again if the analyses I have offered elsewhere (2003; 2007) are correct, it is just part of the very *idea* of *artifactual* kinds (as opposed to biological or chemical kinds) that their natures are fixed at least in part by makers' intentions regarding what features are essential to kind membership—and so ruling out the existence of any kinds with natures of that sort merely begs the question against artifactual kinds.

Let me close by raising one further issue. In her new book (2007), Baker has given us a detailed account of how we can understand artifacts and other everyday objects as constituted by sums of particles, though not identical with them. This is most welcome work, which takes us a good way towards understanding the objects we concern ourselves with in everyday life. But it doesn't cover all artifacts—if we think of artifacts in the broad sense, as the intended products of human labor. For among the artifacts with which we are most concerned are those I've elsewhere (2003b) called "abstract

artifacts"—such everyday objects as novels and laws of state, songs and corporations. While these, like other artifacts, are dependent on human intentionality, such entities as Microsoft, the Patriot Act, or Twinkle Little Star are not themselves constituted by sums of particles at all. In fact, it might be said that our interest is increasingly occupied by abstract artifacts rather than concrete ones, as paper money is replaced with abstract sums in our bank accounts, letters replaced with email messages, and billboards and copies of catalogues with websites. And beyond these replacements, of course, a whole range of new abstract artifacts have come to play central roles in our lives, including computer programs, databases, search engines, and the like. A more thorough account of artifacts must take on this additional project of showing how we can understand these various kinds of abstract artifacts as jointly depending on human intentionality and the physical world, even without being materially constituted at all.

References

Baker, Lynne Rudder. 2008. The Shrinking Difference between Artifacts and Natural Objects. *APA Newsletter on Philosophy and Computers* 07(2):2-5.

Baker, Lynne Rudder. 2007. *The Metaphysics of Everyday Life*. Cambridge: Cambridge University Press.

Elder, Crawford. 1989. Realism, Naturalism and Culturally Generated Kinds. *Philosophical Quarterly* 39:425-44.

Lakoff, George. 1987. *Women, Fire and Dangerous Things*. Chicago: University of Chicago Press.

Thomasson, Amie L. (forthcoming) The Significance of Artifacts for Metaphysics. In *Handbook of the Philosophy of Science Volume: Handbook of Philosophy of the Technological Sciences*, edited by Anthonie Meijers. Elsevier Science.

———. 2007. Artifacts and Human Concepts. In *Creations of the Mind: Essays on Artifacts and their Representation*, edited by Stephen Laurence and Eric Margolis. Oxford: Oxford University Press.

———. 2003a. Realism and Human Kinds. *Philosophy and Phenomenological Research* LXVII(3):580-609.

———. 2003b. Foundations for a Social Ontology. *Protosociology*, "Understanding the Social II: Philosophy of Sociality." 18-19:269-90.

The Shrinkage Factor: Comment on Lynne Rudder Baker's "The Shrinking Difference between Artifacts and Natural Objects"

Beth Preston
University of Georgia

I applaud the direction taken by Lynne Rudder Baker in this fine, short piece. But I think her conclusions are too modest. We can and must go further in the direction she has indicated to ensure that metaphysicians interested in artifacts are finally on the right track after a couple of millenia of errancy.

The main symptom of errancy is the traditional insistence that artifacts are ontologically deficient in comparison to natural objects. Baker argues that none of the traditional ways of picking out substances in fact has this implication. In particular, she says, although artifacts do depend on human intentional states in ways that natural objects do not, this difference does not imply any ontological deficiency in artifacts. Furthermore, she argues, the more general distinction between mind-dependent and mind-independent objects on which the claim of deficiency is often predicated cuts no metaphysical ice. These are fine conclusions. But there are more radical ones in the offing.

Let us start with the general distinction between mind-dependent and mind-independent objects. Baker gives two reasons for regarding this distinction as ontologically nugatory. First, it draws a line in an ontologically unilluminating place in that, for example, it groups insects with galaxies on the mind-independent side, and artifacts with afterimages on the mind-dependent side. Second, advances in technology are increasingly blurring this line anyway by producing objects that are ambiguously natural and artifactual. Although Baker does not put it this way, this second reason provides further support for the claim that the distinction is unilluminating—not only does it assort things oddly in general; it fails to assort some things at all. Moreover, if it should fail to assort a *lot* of things and/or important kinds of things, we would be in a position to draw the stronger conclusion that the distinction cannot be applied reliably across much of the territory it is alleged to partition and is therefore not viable.

However, on Baker's view we are not in this position—at least not yet. Her view is that in a few recent cases the line between the natural and the artifactual has been blurred. As advertised in the title of her article, she predicts that such cases will become more and more common as technology advances, and that the perceived significance of the distinction between mind-dependent and mind-independent objects will fade proportionately. Baker gives four examples to support this claim about blurring (7).

- "Digital organisms" that can reproduce, mutate, and so on, all without any human intervention other than the initial programming effort.
- "Robo-rats" that have electrodes implanted in their brains to "direct" their activity.
- "Bacterial batteries" operating by means of bacteria that naturally produce electrical energy.
- "Search-and-destroy" viruses that are genetically engineered to target cancer cells.

Each of these exemplifies a different way of blurring the difference between artifactual and natural objects. But unfortunately for Baker's view, none of them has anything inherently to do with advances in technology.

Genetically engineered viruses blur the line between artifactual and natural objects because they exemplify human intervention in natural, genetic processes to produce organisms that better serve human purposes. But this makes genetic engineering just the most recent method of domesticating other living organisms. And domestication is a practice as old as the hills and completely ubiquitous. It is now believed to have originated independently in seven different areas across the globe, beginning with the domestication of wheat in the Near East about 10,000 years ago.[1] There are some differences between modern genetic engineering and historical forms of domestication, of course. First, it is widely believed that at first human interventions in genetic processes were unintentional. For example, wild wheat has seed heads that shatter when touched, thus distributing the seeds widely over the ground. Good for the plant; bad for the paleolithic seed gatherer. However, non-shattering heads occur as a relatively frequent mutation. Gatherers would have ended up with relatively more seeds from these than from shattering heads provided they harvested the wheat by cutting it with a sickle or pulling up the plants. Then when they started planting these seeds themselves, they slowly but surely created predominantly non-shattering strains of wheat.[2] Second, until Mendel came along no one had any idea exactly *what* they were intervening in when they intentionally bred preferentially from plants and animals with desired characteristics. And finally, until genetic engineering came along even this intentional intervention was accomplished indirectly by selection of phenotypes rather than directly by manipulation of the genotype. But the relative explicitness of the intention to intervene, the relative

sophistication of the knowledge involved, and the technical means to intervene relatively more directly are superficial and varying differences. What is fundamental and constant is the human intervention. Thus the line between the natural and the artifactual was blurred in this way as soon as domestication began.

Digital organisms blur the line in a similar way. They are virtual entities, but they are explicitly modeled on natural organisms. Indeed, they are often used to study evolutionary processes, since they are much easier to manipulate in controlled ways than naturally occurring organisms in their home environments. Moreover, since digital organisms are used for this and other human purposes, and since their "genetic" processes are modified *ad libitum* to suit these purposes, they are in effect domesticates created from scratch out of non-living material. So, like genetically engineered organisms, they are just a recent, if startling, development in the very long history of domestication.

Let us now consider bacterial batteries. They blur the line between the natural and artifactual by incorporating naturally occurring organisms to perform a specific function as part of an artifact. But this, too, is an ancient practice. A ubiquitous example, the origins of which are lost in the mists of prehistory, is the use of fermentation in brewing and baking. Beer and leavened bread are attested in ancient Egypt *circa* 5,000 years ago, but many historians of food agree that their origins are probably much earlier, perhaps as early as the domestication of cereal grains.[3] In any case, fermentation is a common process in nature and easily coopted for human purposes. Other examples of this kind of blurring include cheese, wine, vinegar, soy sauce, and yoghurt, all also of ancient origin and as ubiquitous previously as now.

Robo-rats are the converse of bacterial batteries. Instead of a naturally occurring organism performing a function as part of an artifact, an artifact performs a function as part of a naturally occurring organism, thus blurring the line in the opposite sort of way. In the robo-rats, three wires are implanted in neurons connected to a rat's right whiskers, left whiskers, and an area that causes pleasurable sensations, respectively. The rat is then trained to go right or left in response to stimuli to the whiskers on the corresponding side by rewarding it with stimulation in the pleasure area. Now this is just ordinary training, so the "directing" of the rat's movements exhibits no novelty.[4] What is new here is only that the stimuli are delivered directly to the brain rather than through the senses. So this is analogous to the modern ability to place a metal pin *in* a bone to hold it together while it heals rather than placing a splint or cast on the outside of the limb. At first blush, it seems this phenomenon would be absent in the earlier stages of human history because of the lack of safe technologies for implanting devices inside the body. But one good example is tattooing, which implants ink into the skin—again a very ancient and widespread practice.[5] More importantly, though, there is a continuum between artifacts that are implanted in the interior of the body, those that are attached to its surface, and those that are manipulated by the person. Consider this series: artificial hippocampus,[6] artificial heart valve, cochlear implant, dentures, artificial arm, rake. A rake extends the capability of hand and arm rather than replacing it. But as Merleau-Ponty points out, for skilled users such artifacts function as parts of the body.

> To get used to a hat, a car or a [blind person's] stick is to be transplanted into them, or conversely, to incorporate them in the bulk of our own body. Habit expresses our power of dilating our being-in-the-world, or changing our existence by appropriating fresh instruments. (Merleau-Ponty 1962, 143)

And this power is undoubtedly even older than the human use of tools, since it is a power enjoyed to some extent by some non-human animals, as well as by now extinct, tool-using hominids. In short, here again we have a way of blurring the line between the natural and the artifactual that is ancient and ubiquitous.

What we must conclude, *pace* Baker, is that the difference between the artifactual and the natural is not shrinking or becoming blurry. It always was blurry. It is not shrinking because there never was an ontological gap between natural objects and artifacts except in the philosophically and theologically heated imaginations of human beings in some cultures. This has unfortunate consequences for Baker's prediction that advances in technology will erode the perceived significance of the distinction between mind-dependent and mind-independent objects. The problem is not that philosophers have had no examples to hand until recently that blurred the difference between the natural and the artifactual. The problem is that even though surrounded on all sides by such examples they have ignored them.[7] So we will have to try some other tack to dispel the perceived significance of the distinction between mind-dependent and mind-independent objects, and with it the temptation to the ontological deficiency thesis.

Our best bet, I think, is to scrap the more restricted version of the distinction between mind-dependent and mind-independent objects Baker continues to accept, *viz*., the distinction between intention-dependent (ID) and intention-independent objects (non-ID) objects. It should be noted that Baker herself does not present this latter distinction as a version of the distinction between mind-dependent and mind-independent objects. But insofar as the intentional states in question are constitutive of a particular kind of mind, this seems like a reasonable interpretation of the relationship between these two distinctions. In any case, Baker accepts the distinction between ID and non-ID objects because she thinks it marks an ontological divide between artifacts and natural objects (6). Artifacts, she thinks, depend ontologically on human intentional states. They exist only because we have certain beliefs and purposes, and they have the proper functions we intend them to have (2-3). Natural objects, on the other hand, would exist no matter what our beliefs and purposes were, and have their proper functions independently of what we might believe or wish those proper functions to be.

On Baker's view, then, the distinction between ID and non-ID objects does have ontological significance insofar as it demarcates artifacts from natural objects. But she vociferously (and rightly, in my opinion) rejects the idea that it is ontologically significant in the sense that it could be used to support the ontological deficiency thesis. She argues, first, that if the criterion for being real is having causal effects, ID artifacts are no less real than non-ID natural objects. Second, she argues, since human beings are part of nature the ontological deficiency of artifacts is really premised on the idea that what is real is only what would exist if there were no human beings; and this is an insufficient basis for that conclusion.[8] I have no quarrel with these arguments. But accepting the distinction between ID and non-ID objects and then trying to limit its influence, as Baker does, leaves the proponent of the ontological deficiency thesis in possession of a foothold. Moreover, there are good reasons for simply abandoning the distinction between ID and non-ID objects.

The first reason goes back to Baker's own claim that some objects are ambiguously natural and artifactual. This means they are ambiguously ID and non-ID—wheat, for example, is the way it is in part because of human practices and in part because of factors completely independent of human beings and their

activities. Consequently, we have to say that the distinction between ID and non-ID objects does not distinguish *neatly* between artifacts and natural objects. Now you do not want to reject an otherwise useful distinction just because of a few, indeterminate cases. But as I have argued above, what we are dealing with here is not just a few such cases, but a lot of them. Moreover, they include whole ranges of historically significant and common kinds of objects—domesticated plants and animals, common foods, prostheses, and skillfully manipulated tools. Perhaps even the human body—are those of us with a lot of ink (or plastic surgery) artifacts or natural objects? We would very much like to know! In any case, the wide range and the importance for human life of such ambiguously ID and non-ID objects suggests that the distinction between artifacts and natural objects is *itself* ontologically unilluminating.[9] There is no sharp divide here, but a smooth continuum. But if there is no good reason to draw a sharp line between artifacts and natural objects, there is *a fortiori* no good reason to retain the distinction between ID and non-ID objects for this ontological purpose. In short, the distinction between ID and non-ID objects is a restricted version of an ontologically unilluminating distinction aimed at explicating another ontologically unilluminating distinction. As such, it is a distinction we do not need and should not want.

Second, it is unclear that the distinction between ID and non-ID objects would help us very much in distinguishing between artifacts and natural objects in any case, because it is itself desperately in need of explication. Moreover, once explicated, it is not clear that it can be used as its proponents propose. This is a very large topic, so I will just give one quick example of the problems involved. As Baker notes, one of the reasons artifacts are typically thought to be ID objects is that their proper functions are held to be dependent on human intentions. First, this assumes an awful lot about the correct account of artifact function. Since there is very little literature specifically on artifact function, it is fair to say that at this point most of the big issues are still up in the air, including the issue of where and how artifacts get their proper functions. More importantly, some of those who *have* studied artifact function specifically, including myself, are disposed to doubt that the proper functions of artifacts *are* dependent on human intentions in any relevant sense.[10] If we are right, it will not be possible to distinguish artifacts from natural kinds by looking for things with intended proper functions. So it appears the distinction between ID and non-ID objects is, again, a distinction we cannot use and should not want. Especially if it secures a foothold for the proponents of the ontological deficiency thesis, which I heartily concur with Baker in rejecting.

Endnotes

1. See Smith 1995, 11-13.
2. Importantly, from a genetic and statistical point of view this process could have been completed in a matter of a few centuries (Smith 1995, 72-74).
3. See http://www.foodtimeline.org/foodbreads.html (accessed May 29, 2008) which incorporates copious scholarly citations.
4. This is pointed out on several websites describing the robo-rats. For example, see http://news.nationalgeographic.com/news/2002/05/0501_020501_roborats.html (accessed May 29, 2008), which also describes the robo-rat project in detail.
5. As just about everybody now knows, Ötzi, the Iceman, who died about 5,000 years ago, had tattoos (see http://en.wikipedia.org/wiki/%C3%96tzi_the_Iceman, accessed May 29, 2008). But only some dots and dashes. Much more elaborate tattoos are known from mummies around the same age from the Tarim basin in what is now China (Mallory and Mair 2000) and from Siberia (Rudenko 1970; also see http://en.wikipedia.org/wiki/Pazyryk).
6. These are not yet available for human beings, but see http://www.newscientist.com/article/dn3488-worlds-first-brain-prosthesis-revealed.html (accessed June 2, 2008).
7. Why is a good question. But it is too big a question to address in this commentary.
8. Interpreted this way, it seems to me Baker's opponents would also have to concede that human beings are not real, since human beings would not exist if there were no human beings any more than artifacts would.
9. A similar conclusion is reached by Dan Sperber (2007). He reaches it by a somewhat different but equally interesting route through consideration of biological and artifactual functions.
10. See Elder (2007) and Preston (2003 and 2006), for example.

References

Elder, Crawford. 2007. On the Place of Artifacts in Ontology. In *Creations of the Mind: Theories of Artifacts and Their Representation*, edited by Eric Margolis and Stephen Laurence. Oxford and New York: Oxford University Press: 33-51.

Mallory, J.P. and Victor H. Mair. 2000. *The Tarim Mummies: Ancient China and the Mystery of the Earliest Peoples from the West*. London: Thames & Hudson.

Merleau-Ponty, Maurice. 1962. *Phenomenology of Perception*, trans. Colin Smith. London and New York: Routledge.

Preston, Beth. 2006. The Case of the Recalcitrant Prototype. In *Doing Things with Things: The Design and Use of Everyday Objects*, edited by Ole Dreier and Alan Costall. Aldershot: Ashgate: 15-27.

Preston, Beth. 2003. Of Marigold Beer – A Reply to Vermaas and Houkes. *The British Journal for the Philosophy of Science* 54:601-12.

Rudenko, S.I. 1970. *Frozen Tombs of Siberia*, trans. M.W. Thompson. Berkeley: University of California Press.

Smith, Bruce D. 1995. *The Emergence of Agriculture*. New York: Scientific American Library.

Sperber, Dan. 2007. Seedless Grapes: Nature and Culture. In *Creations of the Mind: Theories of Artifacts and Their Representation*, edited by Eric Margolis and Stephen Laurence. 124-37. Oxford and New York: Oxford University Press.

Interesting Differences between Artifacts and Natural Objects

Peter Kroes and Pieter E. Vermaas
Delft University of Technology

A review of Lynne Rudder Baker's (2008). "The Shrinking Difference between Artifacts and Natural Objects." American Philosophical Association Newsletter on Philosophy and Computers *7(2): 2-5.*

Lynne Rudder Baker argues in "The Shrinking Difference between Artifacts and Natural Objects" (2008) against the position that the mind-dependency of artifacts makes those artifacts ontologically deficient as compared to natural objects. The argument consists of two parts. First, Baker considers five standard conditions for singling out ontological genuine substances and then reasons that artifacts and natural objects fare equally good or equally bad in meeting these conditions. Second, she challenges the view that the distinction between mind-dependency and mind-independency should be one that serves as a foundation for metaphysics. Baker approaches the topic with her Constitution View (Baker 2000), but the argument she presents is not critically depending on this view—the reasoning is general, clear, and readily accessible for a reader with an appetite for ontology or metaphysics.

The analysis presented by Baker is an important contribution to an emerging trend in metaphysics to give artifacts a proper

position within metaphysics. The orthodox position is that artifacts are indeed ontologically deficient objects. Natural objects like elementary particles as described by the natural sciences are by current orthodoxy the objects that really exist, and artifacts are merely aggregates of those particles that exist qua aggregates of elementary particles but not qua artifacts. This orthodoxy is now challenged by authors like Baker (2004; 2007) and Thomasson (2003; 2007). By this challenge artifacts should also be included in ontological schemes as real objects.

We are very congenial to Baker's position that the mind-dependency of artifacts does not signal any ontological inferiority of artifacts with regard to natural objects. If it would, we are living predominantly in an ontologically inferior world, since our life world is saturated with artifacts. In our own research on technical artifacts, which has its origin in conceptual, methodological, and epistemic analysis of engineering and the engineering sciences, we also take artifacts as ontologically mind-dependent, since they have a dual nature being physical and intentional constructions at the same time (Houkes, Vermaas et al. 2002; Kroes and Meijers 2006). Our research has brought us increasingly nearer to ontological and metaphysical matters, and we are in strong support of the described development to include artifacts qua artifacts in ontological schemes: also technical artifacts as described by engineering should have a proper place next to the objects described by the natural sciences.

Our aim with this comment on Baker's "The Shrinking Difference between Artifacts and Natural Objects" (2008) is twofold. On the one hand we will focus in detail on parts of Baker's reasoning and then present some criticisms. Referring to the work of Wiggins (2001), Baker discusses five ways of characterizing ontologically genuine substances, none of which, she claims, leads to the conclusion that natural objects are ontologically genuine substances and artifacts are not. We will comment on the first (genuine substances have an internal principle of activity), the second (there are laws that apply to genuine substances), and the fifth one (the mind-independency of genuine substances). On the other hand, we will take distance to the particulars of the paper and explore the question of whether Baker's aim is best realized by her reasoning. In this regard we will criticize Baker's overall approach to the ontological upgrading of artifacts. In general there are two strategies available for emancipation: one can present that which should be acknowledged (in an ontology, in our case) as actually already quite similar to that which is already accepted (in that ontology); or one can present what should be acknowledged as making up a separate (ontological) domain that is different from the original (ontology) but as valuable as that original one. Already judging from the title of the paper, Lynne Rudder Baker opts for the ontological emancipation of artifacts by the first strategy by arguing that they are not (that) different from natural objects. We reason that there are good reasons to adopt the alternative strategy by taking artifacts as ontologically quite different from natural objects but not necessarily inferior.

Internal principles of activity

The first characterization of ontologically genuine substances is the Aristotelian one that they have an internal principle of activity. Baker claims that this characterization does not discriminate between artifacts and natural objects because "[a] piece of gold is a natural object, but today, we would not consider a piece of gold…to have an internal principle of change; conversely, a heat-seeking missile is an artifact, but it does have an internal principle of activity" (2008, 3).

These examples and the underlying line of reasoning are in our view rather problematic. Aristotle's idea of an internal principle of activity can easily be reinterpreted such that a piece of gold has an internal principle of activity, namely, the physicochemical laws that determine its properties. Gold has the internal principle of activity to dissolve in *aqua regia* or the internal principle of activity not to react with ordinary water. In contrast to Aristotle's original idea, these principles of activity are no longer teleological in nature, but that does not disqualify them as internal principles of activity. A pebble (heavy object) has the internal principle of motion, when released, to fall according to Galileo's law of free fall; this is not a teleological principle of motion, but nevertheless is an internal principle of motion.

It is, moreover, questionable whether the heat-seeking missile has an internal principle of motion qua artifact. Considered as merely a physical object, albeit a rather complicated one, the missile may be said to have, in line with the above remarks, an internal principle of activity. Its motion is determined by the complex physicochemical processes that are taking place in the missile and is governed by the laws of nature. As a physical object, the missile (or any other kind of human-made physical object) is not different from a pebble or a piece of gold: it is a natural object because it has its own principle of activity. But does it have an internal principle of activity as an artifact, as a heat-seeking missile? That is what Aristotle would deny, because the principle of activity of artifacts lies in the maker of the artifact, not in the artifact itself (a piece of wood of a bed when planted lacks an internal principle of activity to grow into a bed). According to Baker it has, but it is not clear what kind of internal principle of activity she is referring to. What is the internal principle of the heat-seeking missile, qua heat-seeking missile? Is that the principle that it tracks a heat source? If so, to what extent is this an internal principle of activity of the artifact that goes beyond the internal principle of activity it has as a physical object? From an engineering point of view such an internal principle of activity qua artifact appears reducible to the internal principle of activity the object has in so far it is a physical object.

Hence, the first characterization of ontological genuine substances may be discriminative between artifacts and natural objects after all. Under a modern interpretation of internal principle of activity a piece of gold has one; and as long as it is not clear how Baker understands the internal principle of activity of a heat-seeking missile qua missile, it may be maintained that such a missile does not have one.

Let us take distance to the actual argument and switch to a more explorative style. In Baker's Constitution View, "[a]rtifacts…essentially have intended proper functions, bestowed on them by beings with beliefs, desires, and intentions" (2008, 3). We suggest that this proper function may be taken to be its internal principle of activity. Artifacts, when used properly, ought to do certain things; our cars, for instance, ought to transport us from one place to another. Such a position does make sense, turns artifacts, under an additional assumption, into ontological genuine substances, and has moreover a number of surprising consequences. The proper function of a technical artifact can indeed be taken as a principle of activity of the artifact qua artifact because the function of an artifact is not reducible to the physical properties of the artifact (or to its physical principle of activity). As already alluded to in the last quotation of Baker, functions of artifacts depend on features that go beyond the pure physicochemical structure of the artifacts involved. Proper functions depend by virtually all accounts of proper functions on, for instance, the intentions of people and on the way in which the artifact is to be used. The proper function may even be assumed to be an *internal* principle of activity since the function is constitutive for being an

artifact, yet, "internal" cannot not mean intrinsic since as said functions depend also on features beyond the artifact.

According to this line of reasoning, artifacts would have internal principles of activity, just as natural objects, so they would be genuine ontological substances by the first characterization. The only difference with natural objects would be that the internal principles of activity of artifacts are not intrinsic to the artifacts. Yet this difference is not without consequences and may be used to actually identify interesting ontological differences between natural objects and artifacts. For instance, artifacts can come into existence temporarily by physical objects acquiring and losing functions even though these physical objects do not change at all qua physical objects. A pebble that is intentionally picked up from the beach and thrown at a stray dog is temporarily a projectile artifact, at least on some accounts of technical functions (e.g., Neander 1991; McLaughlin 2001), although the pebble need not change its physical properties. Moreover, artifacts can acquire new internal principles of activity in place of or in addition to their existing one when they acquire new proper functions by new uses (Preston 1998; Houkes and Meijers 2006), which is a phenomenon for which there seems to exist no counterpart in the realm of natural objects.

Laws

According to the second characterization objects are genuine substances only if there are laws that apply to them. Baker rejects the position that there are no laws that apply to artifacts because "[e]ngineering schools have courses in materials science (including advanced topics in concrete), traffic engineering, transportation science, computer science—all of which quantify over artifacts" (2008, 3). So there are sciences of artifacts kinds, which means that artifacts are no less genuine substances than natural objects.

Baker is, in our view, right in pointing at the engineering sciences as fields in which classes of technical artifacts are studied and in which engineers try to come up with results that quantify over artifact kinds (such as design rules, relationships between for instance the efficiency of artifacts and design parameters, et cetera). Indeed, libraries of engineering schools are full of books devoted to analyses of artifact kinds. However, her argument is based on the assumption that the kind of regularities that engineers come up with for artifact kinds are similar to the laws pertaining to natural objects. That assumption needs further corroboration. From an epistemological point of view, little is known about the kind of knowledge produced by the engineering sciences, more in particular about the nature of the regularities they come up with—this is a sorely neglected field in epistemology. Many of these regularities pertain to physical/chemical processes that take place in artifacts, and as such these regularities appear to be laws that apply to natural objects and phenomena, and not to artifacts qua artifacts. But what about the regularities that quantify over properties of classes of artifacts? In our opinion there is simply not sufficient evidence to back up the assumption that we are dealing here with laws of the same kind as the laws that pertain to natural objects. So Baker's second argument for artifacts being genuine substances needs further underpinning.

Taking again some distance and assuming that eventually this second argument can be underpinned properly, one can suspect that there is at least one difference between laws that pertain to natural objects and those that apply to artifacts. This difference is that it can be defended that the latter may change over time.[1] Take the case of concrete as mentioned by Baker. It may be argued that the physical substance that constitutes concrete has gradually changed over, say, the last century. Disasters with collapsing bridges and buildings may have had the effect that the rules and regulations constructors have to follow when making and pouring concrete were changed or made more precise. Such a change would have an effect on the laws that pertain to the physical substance that constitutes concrete but also to the way in which it is used. Hence, also the laws applying to concrete may have changed, which probably becomes more plausible by noting that the content of courses taught on concrete has changed over the last century. The obvious explanation of this change in the laws for concrete is that concrete qua physical substance changed. Yet, when assuming that Baker's second argument can be underpinned, then these changing laws apply also to concrete qua artifact. If the argument for changing laws can be made more rigorous (which hinges partly on the question to what extent concrete stays the same technical artifact in spite of changes in its physico-chemical makeup), one again has identified an interesting difference between artifacts and natural objects.

Mind-(in)dependency

Let us turn to the fifth characterization, about genuine ontological substances being mind-independent, which Baker considers being the most interesting one. Baker does not attempt to argue that artifacts may be mind-independent just like natural objects. Rather, she bites the bullet and acknowledges that the mind-dependency of artifacts via their functions does constitute a difference between artifacts and natural objects (see the heading of her second section). But, according to Baker, this difference does not imply that they are ontologically deficient. If we apply Alexander's Dictum—to be real is to have effects—then artifacts are as real as natural objects; they have indeed all kinds of effects on human behavior. Her example of the automobiles, however, contains a curious twist: "[w]hen automobiles were invented, a new kind of thing came into existence: and it changed the world" (2008, 4). The "and" here is strange for to be real and to come into existence, is to have effects, so the argument should run: "when automobiles were invented, a new kind of thing came into existence *because* it changed the world."

Baker concludes her paper with some general remarks on the insignificance of the mind-independence/mind-dependence distinction for the ontological status of artifacts. Apart from the fact that in her opinion this distinction is ontologically not illuminating, she draws attention to the fact that the distinction between natural and artificial objects gets more and more blurred by advances in modern technology. Modern technology creates all kinds of things that are difficult to classify unambiguously as artifacts or natural objects. We agree, but not with some of the (implicit) conclusions she draws from this. First of all, it is not due to modern technology that the distinction between artifacts and natural objects becomes problematic. The moment human beings started to use natural objects found in their environment and to change these objects intentionally, the distinction between artificial and natural objects started to pose problems. How much modification, how much human work (intellectual and physical) is necessary to change a natural object into an artifact? There is no "natural" line to be drawn here, since there appears to be a continuous spectrum of objects ranging from natural objects at one extreme to artifacts at the other.

Although she does not say so explicitly, it seems that Baker takes this as an argument for the insignificance of the distinction between natural objects and artifacts. "Does it matter?" she asks and answers that the distinction will become "increasingly fuzzy; and as it does, the worries about the mind-independent/mind-dependent distinction will fade away" (2008, 4). Hence, the only difference between artifacts and natural objects that she does acknowledge, Baker also brushes away

as insignificant, opening the way to treat artifacts and natural objects ontologically on a par.

In our opinion the distinction between natural objects and artifacts does matter, and with it the mind-independence/mind-dependence distinction. Even if the distinction becomes (or has always been) blurred, there are still clear cut cases where the distinction between natural objects and artifacts makes sense (the Hubble telescope is an artifact "no matter what"[2]). This is not only relevant philosophically (how do we as human beings position ourselves with regard to nature and with regard to the things we make?) but also pragmatically. The distinction between natural and artificial objects has always played a fundamental role in, for instance, patent law (an object can only be patented if it is an invention of a human being, that is, if it is a "mind-dependent" entity and not a natural entity) and will do so in the foreseeable future in spite of the fact that modern technology produces things (such as the Harvard oncomouse) that poses challenging questions, not only for philosophers but also for lawyers, about whether we are dealing here with a natural object or a patentable artifact. Moreover, we have argued in this comment that the mind-dependency of artifacts may lead to all kinds of interesting new phenomena also within the realm of metaphysics itself. Artifacts come into existence and stop to exist by picking up and losing functions even though the physical objects that constitute them do not change at all. Artifacts can acquire new internal principles of activity in addition to their existing one when they acquire new proper functions by new uses. And the laws that pertain to artifact kinds may change over time.

Suppressing such aspects of artifacts seems a high price to pay for allowing them a place in metaphysical schemes. Ontological emancipation of artifacts should, in our view, not be achieved by brushing away interesting differences between artifacts and natural objects. We see no shrinking ontological difference between these kinds of objects. They are interestingly different in that artifacts are mind-dependent entities and natural objects are not. This difference, however, does not make artifacts necessarily ontologically inferior to natural entities. That follows only if the criteria for genuine substances, in particular the criterion of mind-independency, that are suited for objects from the natural sciences, are applied to objects of whatever kind. We are not in the position to provide a knock-out argument against this criterion, but can note that in metaphysics the willingness to dispense with it is gaining ground. Thomasson (2003, 607), for instance, remarks that it may be necessary to seek for a broader picture in order to do justice to the ontological status of "independent parts and aspects of the world, and those that are in part our own construction." What is at stake here is an issue that is of wider significance than the mind-(in)dependency issue, namely, the issue about the adequacy criteria for an ontology. There appears to be no generally accepted list of such criteria. For some an ontology that has no genuine place for most of the objects that we deal with in daily life is inadequate, for others not. Without consensus on these adequacy criteria, differences of opinion about the ontological significance of mind-(in)dependency will be hard to settle.

All in all, we share with Baker the conviction that any adequate ontology should contain artifacts as ontologically respectable inhabitants of our world and it is to her credit that she has put this problem on the philosophical agenda. We disagree with her about the contours of such an ontology; instead of assimilating the ontology of artifacts to the ontology of natural objects, we have no difficulty in allowing ontological differences between these kinds of objects.

Endnotes

1. We are agnostic about whether this difference immediately proves that laws pertaining to natural objects and laws pertaining to artifacts are of essentially different kinds.
2. This claim is to be taken not only as a kind of factual statement, but also as a normative one; however the distinction between natural objects and artifacts is going to be drawn, the Hubble telescope should end up as an artifact.

References

Baker, L.R. 2000. *Persons and bodies: a constitution view*. Cambridge, UK; New York, Cambridge University Press.

Baker, L.R. 2004. The ontology of artifacts. *Philosophical Explorations* 7:99-111.

Baker, L.R. 2007. *The metaphysics of everyday life: an essay in practical realism*. Cambridge, UK; New York, Cambridge University Press.

Baker, L.R. 2008. The Shrinking Difference between Artifacts and Natural Objects. *American Philosophical Association Newsletter on Philosophy and Computers* 7(2):2-5.

Houkes, W. and A. Meijers. 2006. The ontology of artefacts: the hard problem. *Studies in History and Philosophy of Science* 37(1):118-31.

Houkes, W.N., P.E. Vermaas, et al. 2002. Design and use as plans: an action-theoretical account. *Design Studies* 23(3):303-20.

Kroes, P. and A. Meijers. (2006). The dual nature of technical artefacts. *Studies in History and Philosophy of Science* 37:1-4.

McLaughlin, P. 2001. *What Functions Explain: Functional Explanation and Self-Reproducing Systems*. Cambridge, Cambridge University Press.

Neander, K. 1991. The teleological notion of 'function'. *Australasian Journal of Philosophy* 69(4):454-68.

Preston, B. 1998. Why is a wing like a spoon? A pluralist theory of function. *Journal of Philosophy* 95(5):215-54.

Thomasson, A.L. 2003. Realism and human kinds. *Philosophy and Phenomenological Research* 67(3):580-609.

Thomasson, A.L. 2007. Artifacts and human concepts. In *Creations of the mind: essays on artifacts and their representations*, edited by S. Laurence and E. Margolis. 52-73. Oxford, Oxford University Press.

Wiggins, D. 2001. *Sameness and Substance Renewed*. Cambridge, Cambridge University Press.

Book Review

Ordinary Objects

Amie Thomasson (Oxford University Press, 2007).

Reviewed by Huaping Lu-Adler
University of California, Davis

There have been various philosophical attempts to eliminate ordinary objects from ontology, including arguments from causal redundancy, colocation problem, vagueness, composition, rivalry with science, and parsimony. In *Ordinary Objects* Thomasson gives a characteristically *meta*-ontological defense of the "commonsense ontology" (of ordinary objects) against these eliminativist arguments. She invokes two views concerning language as tools in defusing these arguments: that there are analytic entailments among claims, and that there are significant constraints on the answerability (truth-evaluability) of existence or counting questions (claims). Some eliminativist arguments are to be deflated based on the former, and others based on the latter. This linguistic and deflationist approach, it is to be noted from the start, is not meant "to provide linguistic solutions to metaphysical problems, but rather to show that what appear as problems for a particular metaphysical view (the view that there are ordinary objects) are in fact no problems at

all, resulting as they do only from misunderstandings bred in misuses of language" (p. 180).

There are three major mistakes about language that Thomasson recognizes in eliminativist arguments: (1) failing to recognize that many important metaphysical principles are not completely general, inapplicable to cases where there are analytic entailments; (2) failing to realize that the most basic claims about existence, identity, etc. of the objects we refer to are analytic; and (3) mistreating generic existence or counting questions (claims) as answerable (truth-evaluable) (p. 177). The first two are attributed to the arguments from causal redundancy (ch. 1) and colocation problems (ch. 4), the third to the arguments from the special composition question (ch. 7) and rivalry with science (ch. 8). And all three are found in the argument from parsimony (ch. 9). To get a flavor of Thomasson's strategy, let's look at her treatment of the argument from causal redundancy and the argument from the special composition question, respectively.

Thomasson's account of analytic entailment relies on two things: first, terms have meanings that are discoverable by conceptual analysis; second, there are analytic interrelations among these meanings. The first point seems to face the challenge from pure causal theories of reference, which take meanings of terms to be determined purely by causal relations to things in the world and not at all by any concepts that competent speakers associate with them (p. 38). Noting two key problems of such theories (the qua problem and the problem of handling nonexistence claims), Thomasson proposes a hybrid theory of reference according to which the reference of a term is determinately fixed only to the extent that the term is associated with a conceptual content which determines what sort (category) of entity it is to refer if it refers at all. This conceptual content determines the term's "frame-level application conditions" (conditions conceptually relevant to whether or not its reference is established) and its "frame-level coapplication conditions" (conditions under which it would apply again to one and the same referent) (pp. 39-40). Competence with the term requires at least tacit understanding of such conditions. In this connection, analytic entailments may obtain in the following way: for any terms "p" and "q," if the application conditions for "p" are also sufficient conditions for "q" to apply, the claim "p exists" analytically entails the claim "q exists." In general, there is an analytic entailment between two claims just in case a competent speaker (and reasoner) could infer one from the other based solely on knowing the truth of the latter and understanding the meanings of the relevant terms. So construed, analytic entailments can obtain without requiring there to be any synonymy, paraphrasing, or reductive analysis relation between the concerned claims (pp. 44-45). This gives Thomasson's account of analytic entailment significant leverage against, among other things, the Quinean attacks on analyticities based on synonymy (ch. 2).

This notion of analytic entailment is pivotal to Thomasson's attempt to deflate a whole array of eliminativist arguments. Take the argument from causal redundancy, for instance. The basic problem with this argument is, according to Thomasson, that although the causal principle it invokes is a legitimate metaphysical principle, the principle's application is limited. Application of the principle presupposes that the concerned claims to causation are analytically independent (p. 11). In general, the presupposition fails for entities x and y when claims of x's causal relevance analytically entail claims of y's causal relevance. This has an important consequence: if a causal claim φ analytically entails another causal claim Ψ, then Ψ requires no more truth-makers than what is already required by φ; especially, no more causal action in Ψ is required beyond what is required in φ. Thus, there is no doubling of the two causal claims, that is, no overdetermination (p. 16). In that sense, the presupposition fails in the particular cases the eliminativist is interested in. Consider the case where a baseball is (commonly) thought of causing the shattering of a window. The shattering is causally overdetermined only if both the baseball *and* the atoms arranged baseballwise caused it, where the two claims to causation are analytically independent. But they are not: the claim that atoms arranged baseballwise caused the shattering analytically entails that a baseball caused the shattering. There is no real overdetermination, and so recognizing the atoms' causal role does not commit us to denying the baseball's. Therefore, the argument from causal redundancy has failed to eliminate ordinary objects from ontology (pp. 17-19).

A few other eliminativist arguments are to be deflated based on the recognition that there are significant constraints on what sorts of existence (or counting) questions (claims) are meaningful or answerable (truth-evaluable). Typically, in specific questions like "Does N exist?" "Do Ks exist?" and claims like "N exists" and "Ks exist," the name "N" and the kind term "K" are associated with certain categorial terms. Such questions (claims) can be answered (evaluated for truth) in two steps: first, conducting conceptual analysis to fix the relevant categorial terms and determine, in accordance with their frame-level application conditions, what it would take for there to be entities of the relevant kinds. Second, doing the empirical investigation, to discover whether these conditions are fulfilled in the world. There are cases, however, in which general terms like "object" and "thing" are used instead. Certainly they can be handled in the same way as the specific questions (claims) are handled, so long as these terms are used as sortals, associated (by speakers) with particular application conditions that determine what it would take for there to be an object (thing) in a given situation (p. 112). But most metaphysical debates rely, Thomasson observes, on claims about whether there is an object (thing) or how many objects (things) there are in a certain situation which involve a non-sortal, purely generic use of "object" or "thing." According to Thomasson, such generic existence claims are incomplete and not truth-evaluable. Similarly, a generic existence question is only an incomplete, unanswerable pseudo-question, in that "no straightforward answer to it, stated in the same terms as the original question, is truth-evaluable" where this reflects "deficiencies in meaning" and not our epistemic shortcomings (p. 113). Although such a question might be revived by using "object" ("thing") as a "covering term," in that sense of "object," whether there is some object there entirely depends on whether some associated categorial term applies: the application of the latter analytically entails that of the former. So construed, the question "Is there some object (thing) here?" is reduced to the more specific "Is there some C here?" where "C" is a disjunction of categorial terms.

The argument from the special composition question (SCQ) is dismissed precisely for the reason that it treats a generic existence question as answerable in a uniform way. On Thomasson's view, by its very nature the SCQ admits no uniform answer. It is therefore illegitimate for the argument from the SCQ to demand one. In a nutshell,

1. The argument from the SCQ is successful only if there is reason to demand a uniform answer to the SCQ.
2. The SCQ asks whether or not, given certain basic entities, there is some one *thing* composed by those entities.
3. The term "thing" in the SCQ has one of these three uses: (a) generic use, (b) covering use, or (c) sortal use.

4. If "thing" is used generically, then the SCQ is not answerable.
5. If "thing" is used as a covering term, then there can be only a disjunctive answer to the SCQ.
6. If "thing" is used sortally, then there are competing but equally legitimate answers to the SCQ.
7. Therefore, there is no reason to demand a uniform answer to the SCQ.
8. Therefore, the argument from the SCQ is not successful.

The eliminativist may resist Thomasson's rendering of the SCQ and thereby reject the premise (2). Thomasson has made the obvious observation that, when asked (say) "Is there any *thing* on the tray?" we wouldn't know how to answer the question unless specific application conditions are provided for "thing" so that we know what to look for. However, the eliminativist may contend, this is not how the SCQ is intended. Admittedly, "thing" occurs in van Inwagen's verbal formulation of the SCQ. But it is *inessential* to the formulation of the question. For the question is really something like this: Given that there are particles arranged somehow, is there a composite which is composed of exactly those particles? Since nothing hangs on the use of "thing," Thomasson's argument has lost its sting.

I am not sure whether this would be a good response on the eliminativist's behalf. I am more concerned with the application of Thomasson's notion of analytic entailment. The basic examples Thomasson gives of analytic entailment seem quite intuitive: (1) analytically entails (2), and (3) analytically entails (4) (pp. 44-45).

(1) There is a house.

(2) There is a building.

(3) There is a baseball.

(4) There is a lump of stuff.

As we have seen, however, the kind of entailments that Thomasson relies on in deflating some eliminativist arguments proceed from, say, (5) to (6), which do not seem as intuitive.

(5) There exist particles arranged baseballwise.

(6) There exists a baseball.

Apparently, there is a disanalogy between the two sets of examples. The relation between "house" and "building" may be regarded as a *purely* conceptual one. In the taxonomy of concepts, the relation of *house* to *building* resembles that of a species to the genus it belongs to. This warrants the claim that a competent speaker can infer (2) from (1) solely based on her understanding of the terms and knowledge of the truth of (1). It may as well be the case that it is constitutive of a speaker's competence with the terms "house" and "building" that she can infer (2) from (1). In other words, if someone sincerely asks "Now there is a house at the end of the road; but is there a building there?" we will doubt that she really understands the terms. But it doesn't seem to be the case with "baseball" and "particles arranged baseballwise." It is plausible to grant a speaker competence with both terms who nevertheless can't infer (6) from (5) just based on understanding the terms and knowing the truth of (5). To use a Moorean open-question test, such a speaker may sincerely wonder: *Now there are particles arranged baseballwise here; but is there a baseball here?* Do we want to conclude that she simply doesn't understand either "baseball" or "particles arranged baseballwise"? I guess not. It does not seem constitutive of competence with either term that a speaker can infer (6) from (5). Even if someone does make the inference, it may not be based *solely* on understanding the terms and knowing the truth of (5)—some substantive views about (say) part-whole relations are assumed. Given that these kinds of inferences are precisely the ones that the eliminativist is challenging, it would be question-begging (against the eliminativist) to insist that anyone who sincerely hesitates about or questions the inferences is not really competent with the relevant terms. If analytic entailment is to play its role in deflating the eliminativist's arguments without begging questions against the latter, the apparent disanalogy noticed here between the more familiar cases and the cases the eliminativist is concerned with has to be addressed directly.

Whether this is a real challenge or not, Thomasson's account of analytic entailment is attractive in its own right. It offers a genuine, philosophically significant, post-Quinian defense of analyticity. If anything, the above concern about disanalogy suggests a need for further work. There are many other things in Thomasson's book that readers of various research interests will find thought-provoking. For it covers, in a well-informed and profound way, a whole range of semantic and metaphysical topics that have been debated among present-day philosophers. Given Thomasson's clear, careful, and straightforward writing style, moreover, even readers with little relevant background knowledge will find the in-depth discussions rather easy to follow.

Papers on Online Education

Access to Information: The Virtuous and Vicious Circles of Publishing

H.E. Baber
University of San Diego

In Spring 2008 I went textbook-free. I linked all and only the readings for my Contemporary Analytic Philosophy course to the class website, along with powerpoints, handouts, and external links to online resources.

Like most of us who teach Contemporary Analytic Philosophy and other courses where the readings are primarily journal articles, I used to use a textbook anthology. Every year I picked the least-worst anthology. I assigned about a third of the readings in the textbook to justify making students buy it and supplemented the textbook readings with books on library reserve, Xeroxes, and online articles. I was fed up.

Textbook anthologies once served an important purpose. Currently, however, most do not facilitate access to information and are not cost-effective. The same is true of hardcopy journals. Initially journals democratized the Republic of Letters. They made information that had previously circulated amongst a small coterie of scholars through private correspondence available to a wider audience. Now Web publishing is cheap and efficient: researchers can make their work available without the help of journal publishers.

Traditional publishing is not outdated and never will be. The book as we know it is a very efficient vehicle for conveying information. Codices knocked out scrolls in the way that quartz watches superceded mechanical watches and CDs replaced records. But Kindle will never knock out traditional books and the Internet will never replace magazines or newspapers. For most purposes, hardcopy books, magazines, and newspapers add value and are preferred by consumers. For some purposes, however, hardcopy publications are not efficient and will likely, in the end, go the way of the scroll, the mechanical watch, and the vinyl record.

There will never be another hardcopy encyclopedia of philosophy like the massive multivolume set published in 1967. The Stanford Encyclopedia of Philosophy and other online resources are cheaper and immeasurably better. Likewise, I shall suggest, textbook anthologies and hardcopy journals are obsolete.

The End of the Textbook Anthology?

To see why textbook anthologies are inefficient we need to consider what they offer. Minimally these products provide access to primary readings, selection, and organization. Some provide various pedagogical extras including editorial introductions and comments, selected bibliographies, "study questions" and the like. Most are packaged in an aesthetically pleasing format. None of these things are worth paying for.

Access to primary readings for most courses we teach is unproblematic: most readings are readily available online and those that are not can be scanned and put up at class websites or online library reserve. Librarians and bookstore personnel, who are knowledgeable about copyright regulations, can help instructors meet legal requirements, which in many cases, can be satisfied by simply password-protecting access. We do not need textbooks to make the readings readily available to students. Moreover, most of us do not need, or want, the selection and structure that textbooks provide. We are as qualified as textbook editors to select readings for our courses and organize them by topic, and much better situated to tailor our selections to suit our interests and meet our students' needs. The "ancillaries" publishers imagine will attract us are useless or worse. As for aesthetics, admittedly textbooks are more attractive than the three-ring binders full of printouts that students in textbook-free courses produce. But I do not think that such packaging is worth the price of the book or, more importantly, the cost of selecting readings and organizing courses to fit the textbook in order to justify making students buy it.

In some circumstances a textbook is a quick and dirty solution. If we are teaching general education courses on topics in which we have no expertise and little interest, a textbook anthology with the standard articles suitably organized cuts preparation time. However, even if we want the selection and structure textbooks provide, we can get it without buying the book: we can use the table of contents to structure our courses, and link the readings. It is, of course, easier and more convenient to buy the book and pass the costs onto students—but not by much.

It does seem like cheating to appropriate a table of contents without buying the book. But here we ought to ask why. What if we all did it? What if we simply grabbed the tables of contents of textbook anthologies, put them up at our class websites, and linked online readings to the entries?

This would wipe out one of publishers' most popular product lines, making it more difficult for them to operate profitably and so more difficult for them to…produce more textbook anthologies. More poignantly, it would cut down on our publication opportunities. Textbook anthologies provide vita entries and occasionally royalties. Moreover, for every textbook anthology there is one, or more, of our colleagues who toiled to put the thing together—wading through the literature, making the selection and creating the structure, writing introductions and study questions, assembling the project and querying publishers. We would be stealing the fruits of our colleagues' labor, much of it pretty miserable drudge work at that.

But is all this drudgery worth it? There are hundreds of textbook anthologies on the market, which cost thousands of man-hours to produce. The opportunity costs are real: these are hours their editors could have spent working with students, preparing classes, and, of course, doing original research. The selections these books include overlap substantially and most of the work is further wasted because the most important product that they provide, information that was once otherwise inaccessible, is now available on the Internet.

In the past, textbooks and journals provided a medium that increased the amount of information available to students and faculty, who in turn financed publishers so they could make more information available. That was the virtuous circle of publishing. Currently the Internet is a much more efficient medium for disseminating the information that journals and textbooks have traditionally provided so, in an attempt to remain competitive, publishers trick out textbooks with worthless "ancillaries" and make them fatter, glossier, and more expensive to add value (as they see it), restrict online access to the content of journals, sell rights, charge licensing fees, and sue for violations of copyright. This is the virtuous circle turned vicious: in the interests of remaining profitable, publishers attempt to restrict access to information.[1] And that is both wasteful and futile, because information is a *public good*.

Virtuous and Vicious Circles

As a "public good," information is non-rival and non-excludable. It is non-rival: the consumption of information by one individual does not reduce the amount of information available for consumption by others. Currently, given virtually universal access to the Internet, it is also de facto non-excludible: no one can be effectively excluded from consuming it.

Public goods are a well-known problem for market-based systems. The story is familiar: without incentives these goods will not be produced and that is, as economist John Quiggin notes, the rationale for copyright:

> Copyright matters because it provides an economic incentive for authors to create socially valuable content in circumstances where, if they weren't given this incentive, they would do something else. The copyright system is necessary to encourage the creation and use of socially valuable content, or so goes the standard utilitarian justification for copyright.[2]

According to the standard story, without the incentives copyright provides for producers and vendors of intellectual property, consumers would have less access to creative works than they would if there were no restrictions on access because there would be less intellectual property produced. When the market works, copyright and other restrictions on access to intellectual property produce a net gain in access to information.

But sometimes the market does not work and the virtuous circle turns vicious. To see this consider "one of those counterfactuals." As a thought experiment, imagine a worst-case scenario at a possible world where there are no textbook anthologies:

You have emerged from grad school without ever having taken an ethics course and at your first job you are asked to teach "Contemporary Moral Issues."[3] What to do? You Google around and pull up a dozen or so syllabi for Contemporary Moral Issues classes that are being taught by colleagues at respectable universities. You note that there is a shortlist of topics they all do as well as some extras. You quickly learn the basic format for an applied ethics course and start putting together your syllabus using a colleague's syllabus as a model. You set up the structure of topics. (Let's see: some general stuff about utilitarianism and other theories with readings from Rawls, Nozick, and Peter Singer; then abortion, euthanasia, the environment, and so on—gotta use that Judith Jarvis Thompson article on abortion;

maybe some extras, like copyright.) Then you plug in the readings. You include the "classic" articles that appear on all syllabi and check out the others that are conveniently linked, picking what you like.

You are a free rider! (You just learnt that term.) You've gotten the selection and structure for an applied ethics course, which your colleague toiled to create, for free!

But is this a bad thing? It's no skin off of your colleague's nose if you tweek and use his syllabus: the selection of readings and structure of his course is a public good—using them doesn't use them up or in any way detract from their value to him or his students. Of course, with lots of free riders like you around, he can't *sell* that reading list: that's why there aren't any applied ethics anthologies at this possible world. But even without that incentive, he will still create and improve his syllabi because he's got a course to teach, and will still put them up at his class websites for his students' convenience and his own. Widespread free-riding does not diminish the incentives for producing syllabi: it only eliminates the incentives for publishing them in the form of textbook anthologies. In general, as Quiggin points out, "the copyright system does not provide incentives to authors to create valuable content so much as it provides incentives to the intermediaries who guarantee the circulation of this content."[4]

With access to the Internet, and a wide range of syllabi and readings available online, you don't need those intermediaries and, indeed, you and your students are better off without them. Putting together your course in this way means building on the expertise and experience of colleagues, tweeking and improving their materials, and learning, which is surely conducive to good teaching. In fact *everyone* is better off: putting syllabi up at a website and linking readings is much easier, less expensive, and less time-consuming than assembling and publishing a textbook; accessing readings online is cheaper and more convenient for students than buying a textbook and hauling it around. As for the "intermediaries," instead of wasting their time trying to compete with the Internet by bloating textbooks, they are more responsive to consumer preferences and produce more affordable materials.[5]

If this is correct then the restrictions on access to information that create a demand for textbook anthologies are counterproductive. They are costly and do not create any additional incentives for producing information. They perpetuate a vicious circle in which academics do unnecessary menial work and publishers have no incentive to improve the efficiency of their operations. There is, however, an even more vicious circle revolving around the hard-copy academic journal, which has, largely in virtue of academics' professional interest in positional goods, succeeded in beating the market.

Journals

In the past, the hardcopy journal was a vital component of the virtuous circle of publishing—indeed, it kicked research into an upward spiral. Academics produced articles and journals made them available to a wide audience of consumers, who were themselves producers. The more information that was available, the more research was produced: journals proliferated and made yet more research available to a wider audience of academics who were engaged in research and published the results of their research in journals. Life was good.

The hardcopy journal was not, however, an ideal medium, particularly with growing specialization. No one read all the articles in any given issue of any journal and everyone needed to read a dozen or more journals to keep up with work in their fields. Individual subscriptions to journals became largely pointless, unless you could afford to subscribe to a dozen or more. And if you had to go to the library to read journals and Xerox the articles you needed, there was no point in subscribing to *any* journals yourself: you were going to be working in the library anyway. With increased specialization and the proliferation of journals we were regressing to the age of the chained book.

What academics needed was a way to select only articles that were relevant to the areas in which they were working. And the medium that satisfied this need was the Internet. Most articles are available somewhere on the Internet: at their authors' websites, through various pre-print archives, or, with restrictions, in online databases like EBSCO to which academic libraries subscribe. On the Internet, we can search for articles in our areas of interest through the *Philosopher's Index* or simply by Googling; we can collect bibliography, browse current journals, skim articles, and read those that are of interest; and we can work 24/7 from almost anyplace on earth. Life is *very* good.

We do not need hardcopy journals. We do, however, need surrogates that satisfy their selection and credentialing functions. To stay in the game, we need to read articles that are not only of high quality but which other people in our field are reading. Publication in an academically respectable journal signals that an article is worth reading and that other people are reading it. In addition, to get jobs, a scarce resource, and to keep them, we need to accumulate positional goods, in particular, journal publications. The Internet may be the most efficient medium for "publishing," that is, making our work public, but self-publishing on the Internet is professionally worthless because anyone can do it.

Currently, the purpose of journals is not publishing but screening and credentialing. These services are vital because time and jobs are scarce. With limited time, we need to know which articles are worth reading and, since jobs are a scarce resource, we need refereed publications to get and keep jobs. But we don't need paper to meet these needs. It is possible in principle for online facilities to provide those services. *The Philosopher's Imprint*, a free, refereed, online journal published by the University of Michigan Digital Library is the model of what journals should, and one hopes, will become. Arguably, the program described in its mission statement is what we should promote:

> There is a possible future in which academic libraries no longer spend millions of dollars purchasing, binding, housing, and repairing printed journals, because they have assumed the role of publishers, cooperatively disseminating the results of academic research for free, via the Internet. Each library could bear the cost of publishing some of the world's scholarly output, since it would be spared the cost of buying its own copy of any scholarship published in this way. The results of academic research would then be available without cost to all users of the Internet, including students and teachers in developing countries, as well as members of the general public.
>
> These developments would not spell the end of the printed book or the bricks-and-mortar library. On the contrary, academic libraries would finally be able to reverse the steep decline in their rate of acquiring books (which fell 25% from 1986 to 1996), because they would no longer be burdened with the steeply rising cost of journals (which increased 66% in the same period).[6]

The mission statement, however, continues, poignantly: "The problem is that we don't know how to get to that future

from here, and there are so many other, less desirable futures in which we might end up instead."

The problem of getting there from here is exacerbated because the role of traditional journals as credentialing agencies locks in a suboptimal equilibrium. *Ceteris paribus* I might prefer to publish in a free, online journal and wish that everyone else did too. But if I have an interest in professional advancement, and if I want to read articles in my field that others are reading and publish in a place where my article will be read, I will read and publish in traditional journals. I will do that because, as a rational chooser, I know that my colleagues are thinking the same way, and that they will therefore publish in traditional journals, read traditional journals, and assess my professional merits on the basis of publications in traditional journals. We might all wish that things were otherwise, but it will be very difficult to break that vicious circle.

The most feasible way to get there from here I suspect would be for traditional journals to morph into online journals on the model of the *Philosophers' Imprint*—"edited by philosophers, published by librarians and free to readers of the Web." That is, however, not what is happening. Instead, journals increasingly are relying on commercial firms, which make their living by restricting access to journal articles, to manage their Internet affairs. I have just signed away copyright on an article to one of these firms because keeping copyright to enable open access at the site it maintains for that journal would have cost me $3,000.[7] This is the less than desirable future, which, at least in the short run, seems most likely unless we find some way to achieve a more desirable one.

Getting There From Here

The vicious circles I have described persist because we in the profession, in the various roles we play, are not making use of appropriate technology. We dread the start-up costs of using new technologies, overestimate the difficulty of projects as quick and easy as putting up class websites, and underestimate the importance of making our teaching materials and papers available online. We aren't aware of the resources that are available and even where we are blessed with well-funded IT departments don't know what to ask for. More often than not we end up in the classic predicament: we know what we need but don't understand the technology; IT staff understand the technology but don't know what we need; and administrators who neither know what we need nor understand the technology make the purchasing decisions.

We have the resources to get to a better there from here. Within our universities we can collaborate with colleagues, librarians, and IT personnel to facilitate the use of existing and emerging technologies in support of research and teaching. On the Web, the Open Access News[8] provides information about the open access movement devoted to putting peer-reviewed scholarly literature on the Internet, making it available free of charge, and removing barriers to serious research. Sites like MIT Open Courseware[9] and Carnegie-Mellon's Open Learning Initiative[10] are models for the effective use of online resources for teaching. And, within our profession, the APA Philosophy and Computers Committee publishes the current newsletter, organizes sessions at APA meetings, including the one in which an earlier version of this paper was presented, and other projects to support the use of technology in research and teaching in order to facilitate our progress to a future at the best of all accessible possible worlds.

Endnotes

1. http://insidehighered.com/news/2008/04/17/gsu.
2. John Quiggin and Dan Hunter. "Money Ruins Everything." Hastings Communications and Entertainment Law Journal (forthcoming). Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1126088.
3. It happened to me.
4. Quiggin and Hunter, Op. Cit.
5. Affordability is a significant concern at community colleges and other institutions that cater for economically disadvantaged students. See, e.g. http://insidehighered.com/news/2008/04/29/textbooks and http://insidehighered.com/news/2008/04/16/textbooks.
6. http://www.philosophersimprint.org/about.html
7. My contract reads: "if you want your article to be available to everyone, wherever they are, whether they subscribe or not, then you should publish with Open Access. [name of firm] operates a program called Open Choice that offers authors the option of having their articles published with Open Access in exchange for an article processing fee. The standard fee is US$3000. [sic!]If you want to order Open Access, please click the button 'Yes, I order Open Access' below."

   My contract (if I understand it correctly) allows me to self-publish a version of my article at my own website, under various conditions with restrictions, and that is what I will do. What are they playing at? I suppose they imagine that someone might want to reprint my paper, and want to make sure that they can soak him for permissions. But I can't imagine why anyone would want to reprint my paper since it will be up at my website where everyone can get it for free.
8. http://www.earlham.edu/~peters/fos/fosblog.html.
9. http://ocw.mit.edu/OcwWeb/web/home/home/index.htm.
10. https://oli.web.cmu.edu/jcourse/webui/free.do.

What a Course on Philosophy of Computing Is Not

Vincent C. Müller
American College of Thessaloniki

1. Learning goals

This programmatic paper is trying to contribute to the development of an international course in the philosophy of computing, the main outlines of which were discussed at NA-CAP 2007 in Chicago (and at earlier CAP meetings). The chair of the 2007 panel, Piotr Bołtuć, invited the panelists, saying: "The aim is to try to define some standards of what a good course in this would look like," so I will try to contribute to this aim.

A good course should include the interesting issues and anyone in the field can name many such issues, but why this thematic and pedagogical *unity*? Is there anything that holds the course together? This seems the crucial question. It appears that the course cannot be called just "Computing *and* Philosophy" as the CAP conference series and the associations because this title is chosen for its all-inclusiveness of "something to do with both computing and philosophy." Inclusiveness is useful for academic organizations and conferences, but in a course it needs to be avoided. We cannot present a ragbag of interesting issues, but neither can we go around and claim ownership of some established problems, perhaps after giving some of them new names. To put it positively, we must formulate the core of the course, the central learning goals, and then we can see what fits and what does not.

Such learning goals are not already specified by saying that the course is "about the philosophy of computing," not just because this is too vague, but primarily because it does not specify any philosophical problem. The idea that one can take any object, event, or phenomenon x and then do a "philosophy of x" is one that gives philosophy a bad name.

There is a good reason why there is a philosophy of physics, but hardly one of chemistry, why there is a philosophy of history, but hardly one of archeology. The former have interesting and basic philosophical problems, problems that are urgent for the practitioners of these disciplines, the latter do not. If we want to do and teach a philosophy of computing, we need to specify such basic problems that are intellectually challenging and practically relevant.

In the following, I shall take a look at some possible candidates that may provide unity—and essentially reject them all for being both too wide and too narrow. This should result in some controversy and thus in steps forward. The negative results also allow an indication where the positive result should be located.

2. Not philosophy of computer science

The obvious candidate for identification is the fairly well-established field of philosophy of computer science—which depends on the large and wealthy mother discipline. This candidate is too narrow a specification for the course, however, because it is limited to computation in artifacts, in "computers." This ignores a purely formal theory of computing as well as the important possibility of computing in natural processes. It is now common to think of the human mind as operating mechanically, even as operating as a computational mechanism; the view called "computationalism," which was the original myth of the cognitive sciences. (The history of which can now be read in Margaret Boden's opus magnum (Boden 2006); see also my review (Müller 2008).) What is more, there are countless other natural (and even social) mechanisms that can usefully be explained as being computational.

The candidate is also too wide because computer science includes a few problems that are irrelevant to this course, namely, those to do with technical problems of the hardware—problems in electrical engineering, robotics, etc.

Finally, it is far from clear that computer science really has interesting basic problems that are in need for philosophical analysis. In fact, "computer science" is a misnomer since the aim of the discipline is to *make* things, not to *find out* about the world. It is thus a proud member of the *engineering* disciplines, not of the sciences. This detail is often overlooked by all-encompassing representatives of the discipline, like Herb Simon, one of its founding fathers who used the characteristic title "Artificial Intelligence: an Empirical Science" (Simon 1995), arguing that the aim of AI is to find out what intelligence is.

Despite this rejection, I think we should keep in mind a few questions of this area that should feature in a course on Philosophy of Computing, for example: What is software? What is a computer program?

3. Not theoretical computer science

There is a theoretical area that may have been missed in the previous section: the purely theoretical study of computational structures and processes, often considered to be a branch of mathematics.

It should be obvious that this candidate is too narrow, given that there is more to computing than formal structure. There is also a certain danger that it may be too wide, in its discussion of some purely formal subjects and methods that lead in to the wider areas of logic and mathematics.

Having said that, this area is clearly rich in subjects that should feature in a course like ours, questions related to the formal description of computing and computing machines, the bounds of computability, the resources needed (complexity theory), program verification, data structures, etc.

4. Not philosophy of information

One other candidate with reasonably serious aspirations of complete coverage is the philosophy of information—after all, what is called "computer science" in English is often called the equivalent of "informatics" in other languages.

Whether this proposal is too narrow is actually a difficult question because it would presume that we can say whether all computing deals with information. I tend to think that this is not the case, since some computing is purely syntactic; but clearly this is an open question. At any rate, we should not pretend that it has been answered—which is precisely what we would do if we were to limit our subject to the philosophy of information.

The candidate is clearly too wide since not all information is digital, but even if we include non-digital information, not all information is processed computationally, or processed at all, for that matter.

Again, this area is a rich source of relevant questions: Does computation compute over information or just over data? What is information and what is its dynamics? Is intelligence information processing, or could it be? Is nature somehow informational?—Of course, a rich source of programmatic material on these questions is (Floridi 2004).

5. Not philosophy of AI

The philosophy of artificial intelligence is clearly a rich source of basic philosophical problems, but equally clearly it is narrower in scope than our course: there is more to computing than intelligence. This area is also too wide because many of its problems are not specific to computational intelligence, particularly those traditionally discussed in the philosophy of mind and cognition.

The problems that we should take into account from this candidate are the classical ones concerning cognition, perception, and intelligence in computational systems. Particularly important are, of course, the problems of representation or meaning in computers and perhaps the more recent questions whether an embodiment, a will, and emotions are necessary for intelligence in computers.

6. Not philosophy of technology

One area that has been somewhat on the sidelines of philosophical debates within the strongly analytic traditions of the philosophy of computing is the embedding of computation in the wider context of human technologies. As in the philosophy of computer science, however, this area is too narrow, since it ignores computational systems that are not artifacts (e.g., formal, natural, social).

That this candidate is too wide is evident, since it covers any kind of human technology, not just the computational ones. (And if we shall discover that everything is computing, the philosophy of computing would not become all of philosophy.)

This area does, however, provide us with a very important set of questions concerning human life with computers, especially human societies and computers. It also reminds us that there is the important issue of human ethical behavior with computers, what is known as "computing ethics." (The idea of ethics *for* computers or other technical artifacts is just a misunderstanding, in my opinion.)

7. Conclusion: The Core Curriculum

We have seen that each of the above proposals was both too narrow and too wide. In a sense, this is good news, since it shows that the philosophy of computing is neither a subdivision of a traditional field nor a superset of already existing fields. It does thus stand a chance to carve out a field of its

own. It seems essentially situated within the philosophies of information and technology, while incorporating most of the philosophy of computer science and of AI; plus an appendix of applied ethics.

On a positive note, I think that what we have seen so far can be captured fairly concisely to be about theoretical questions on the nature of computing, more practical questions of what that kind of mechanism is capable of, and finally the ethical question. In his model course, Bill Rapaport suggested that "…an excellent course on the philosophy of computer science could consist solely of close readings of Turing's two major essays: his 1936 paper on computability and his 1950 paper on whether computers can think" (Rapaport 2005). I would agree, just that we would need to add a paper on ethics, something like Jim Moor's "What is Computer Ethics?" (Moor 1985).

Immanuel Kant famously defined philosophy to be about three questions: "What can I know? What should I do? What can I hope for?" (KrV, B833). I want to suggest that the three questions of our course on the philosophy of computing are: What is computing? What should we do with computing? What could computing do?

References

Boden, M.A. 2006. *Mind as Machine: A History of Cognitive Science*. 2 vols. Oxford: Oxford University Press.

Floridi, L. 2004. Open Problems in the Philosophy of Information. *Metaphilosophy* 35(4):554-82.

Moor, J. 1985. What Is Computer Ethics? *Metaphilosophy* 16(4):266-75.

Müller, V.C. 2008. Review of Margaret Boden 'Mind as Machine' (2 vols., Oxford University Press 2006). *Minds and Machines* 18(1):121-25.

Rapaport, W. 2005. *Philosophy of Computer Science: An Introductory Course* [accessed 12.3.2007 2007]. Available from http://www.cse.buffalo.edu/~rapaport/

Simon, H. 1995. Artificial Intelligence: An Empirical Science. *Artificial Intelligence* 77:95-127.

Computing and Philosophy Global Course: What can we hope for [from computing]? What should we do [with computing]? What can we know [about computing and by computing]?

Gordana Dodig-Crnkovic
Mälardalen University, Västerås, Sweden

The first Computing and Philosophy Global Course is planned for fall 2008 as a result of collaboration between several European and American universities and with ambition to grow in the future into an even bigger course including more countries worldwide (CaP 2008). The course is based on an earlier Swedish National Course (see Dodig-Crnkovic & Crnkovic 2007). The co-organizers and invited speakers include Peter Boltuc, Keith Miller, Gaetano Lanzarone, Vincent Müller, and Gordana Dodig-Crnkovic, the course coordinator, with involvement of students from respective institutions. Luciano Floridi, Marvin Croy, and Bill Rapaport are associated with the project.

Before describing the course (how?), let me present our motivation for organizing it (why?). Preparing Computing and Philosophy Global Course we have to answer number of questions. Let us take from where Vincent C. Müller's article in this issue "*What a course on philosophy of computing is not*" left us:

"Immanuel Kant famously defined philosophy to be about three questions: "What can I know? What should I do? What can I hope for?" (KrV, B833). I want to suggest that the three questions of our course on the philosophy of computing are: What is computing? What should we do with computing? What could computing do?"

Indeed, those are precisely the questions we will have to answer. I would only broaden the scope and re-phrase them in the following order:

- What can we hope for [from computing]?
- What should we do [with computing]?
- What can we know [about computing and by computing]?

In what follows I will try to answer the above questions, one by one.

What can we hope for [from computing]?

This first question is about the goals of the course. In what way can the course be important and reflect the current and possible future interests of the communities we are supporting? How can we contribute to the development of the field?

The initial step is to understand the state of the art. I take *computing to encompass both computation and information*. As argued in Dodig-Crnkovic (2003), the German, French, and Italian languages use the respective terms "Informatik," "Informatica," and "Informatique" (Informatics in English) to denote Computing. It is worthwhile to observe that the English term "Computing" is empirical, while the corresponding German, French, and Italian term "Informatics" has an abstract orientation. This difference in terminology may be traced back to the tradition of nineteenth-century British empiricism and to continental abstraction, respectively.

The following list[1] of research topics (research presented at Computing and Philosophy (CAP) conferences) illustrates the present day state of the art of the research field:

1. Philosophy of information
   a. Philosophy of information technology (global information infrastructures: technological architectures, converging information technologies, etc.)
2. Philosophy of computation
   a. Philosophical aspects of Bioinformatics, Biocomputation
   b. Computational evolution, Artificial life
3. Computational approaches to the problem of mind
   a. Philosophical questions of Cognitive Science
4. Philosophy of Computing
   a. Philosophy of CS
      i. Models of Logic Software
   b. Philosophy of AI
   c. Computational Linguistics
   d. Philosophy of computing technology
5. Real and virtual, modeling, simulations, emulations
6. Computing and Information Ethics
   a. Roboethics
   b. Norms and Agents
7. Societal aspects of computing and IT
   a. Cultural Diversity and Technoscience Studies
8. Philosophy of Complexity (distributed processes, emergent properties, etc.)
9. Computational metaphysics
   a. Computational ontologies
   b. Computational cosmologies (e.g., pancomputationalism, digital physics)

10. Computational Epistemology
11. Computer-based Learning and Teaching
    a. Distance Learning in Philosophy and Computing

From the list above it is evident that under CAP many research traditions co-exist and as Müller correctly points out, CAP is **not** any of the following: philosophy of computer science, theoretical computer science, philosophy of information, philosophy of AI, or philosophy of technology. It is a forum for cross-disciplinary – inter-disciplinary – multi-disciplinary process of knowledge exchange and establishment of relationships between existing knowledge fields, among others those just mentioned.

At present we are witnessing a major scientific, technological, and global-scale societal transformation that accompanies the extensive use of information networks and computing capabilities in all spheres of knowledge creation. The Computing and Philosophy (CaP) global course will offer a glimpse of a new complexly networked and dynamic world, emerging from the research results in sciences, humanities, technologies, and variety of supporting information-intense fields. This development of a new body of knowledge is followed by a distinct paradigm shift in the knowledge production mechanisms (Dodig-Crnkovic 2003).

Globalization, information networking, pluralism, and diversity expressed in the cross-disciplinary research in a complex web of worldwide knowledge generation are phenomena that need to be addressed on a high level of abstraction, which is offered by philosophical discourse. Examples of philosophical approaches closely connected to the on-going paradigm shift may be found in Floridi (2004 and 2005), Wolfram (2003), Mainzer (2003 and 2004), Chaitin (2005), Lloyd (2006), and Zuse (1967).

The objective of the CaP course is to present philosophical reflection over computing and related phenomena and to provide philosophically interesting insights into current state of the art knowledge in computing and information. We hope to increase the understanding between computing and philosophy by building conceptual bridges between the fields.

In order to understand various important facets of ongoing info-computational turn and to be able to develop knowledge and technologies, a dialogue and research on different aspects of computational and informational phenomena are central. Taking information as a fundamental structure and computation as information processing (information dynamics) one can see the two as complementary, mutually defining phenomena. No information is possible without computation (information dynamics), and no computation without information (Dodig-Crnkovic 2005, Dodig-Crnkovic and Stuart 2007).

What should we do [with computing]? Knowledge as complex informational architecture: Necessity of a multidisciplinary dialogue

Why is it important to develop Computing and philosophy as a multi-disciplinary discourse? One of the reasons is epistemological—it provides the fundamental framework suitable for common understanding of an impressive number of presently disparate fields. This argument builds on a view of knowledge as structured informational construction. According to Stonier (1997), data is a series of disconnected facts and observations, which is converted into information by analyzing, cross-referring, selecting, sorting, and summarizing the data. Patterns of information, in turn, can be worked up into knowledge which consists of an organized body of information. This constructivist view emphasizes two important facts:

- going from data to information to knowledge involves, at each step, an input of work, and
- at each step, this input of work leads to an increase in organization, thereby producing a hierarchy of organization.

Research into complex phenomena (Mainzer 2004) has led to an insight that research problems have many different facets which may be approached differently at different levels of abstraction and that every knowledge field has a specific domain of validity. This new understanding of a multidimensional many-layered knowledge space of phenomena have among others resulted in an ecumenical conclusion of science wars by recognition of the necessity of an inclusive and complex knowledge architecture which recognizes importance of a variety of approaches and types of knowledge (see, for example, Smith and Jenks 2006). Based on sources in philosophy, sociology, complexity theory, systems theory, cognitive science, evolutionary biology, and fuzzy logic, Smith and Jenks present a new interdisciplinary perspective on the self-organizing complex structures. They analyze the relationship between the process of self-organization and its environment/ecology. Two central factors are the role of information in the formation of complex structure and the development of topologies of possible outcome spaces. The authors argue for a continuous development from emergent complex orders in physical systems to cognitive capacity of living organisms to complex structures of human thought and to cultures. This is a new understanding of unity of interdisciplinary knowledge, unity in structured diversity, also found in Mainzer (2004).

In a complex informational architecture of knowledge, logic, mathematics, quantum mechanics, thermodynamics, chaos theory, cosmology, complexity, the origin of life, evolution, cognition, adaptive systems, intelligence, consciousness, societies of minds,[2] and their production of knowledge and other artifacts…all have two basic phenomena in common: information and computation. In the Computing and philosophy global course we will use computing and information as a means to provide a framework for those jigsaw puzzle pieces of knowledge to put together into a complex and dynamic info-computational view.

What can we know [about computing and by computing]?

The main textbook is *The Blackwell Guide to the Philosophy of Computing and Information* (Blackwell Philosophy Guides, 2004), edited by Luciano Floridi.

Following fields will be covered by the CaP global course (CaP 2008):

- Philosophy of Information

The course will give an introduction to Luciano Floridi's Philosophy of Information, including Information Ethics and among others introduce ideas of "infosphere" and "being as being informed" (Floridi 2004-2007). We also introduce philosophy of the web (see Halpin 2007).

- Philosophy of Computation

Computation may be understood as information processing. At present, research on computation is intensely developing new views of the phenomena, especially natural computation (MacLennan 2004, Siegelman 1999), which uses natural phenomena as computing devices. Some relevant questions are: What is computation? How do computation and information relate? Turing machine model vs. interactive computation as closed system vs. open system (Wegner 1998, Goldin 2005, Goldin et al. 2006). Church-Turing thesis domain. Digital vs. analog (Müller 2007). Natural computation

as interactive computation in the world goes in an important sense beyond Turing paradigm. It also calls for new logical approaches (Dodig-Crnkovic 2005 and 2008) and references therein - Abramsky (2003), Allo (2007), Benthem (2006), Hintikka (1973), Japaridze (2007), Kelly (2004), Priest and Tanaka (2004). Pancomputationalism views the whole of the universe as a network of computational processes. Taking information as a structure and computation as its dynamics, info-computationalism is a flavor of pancomputationalism which not only sees computational universe as a process but also as an informational structure. Conceptually, the ambition of new info-computationalism is to explore the possibilities of the real world as a resource of computational devices. In this view the Turing machine computational model is a subset of a more general natural computation.

- Philosophy of Mind

Here we present Computationalism and its critics. Epistemology naturalized (Dodig-Crnkovic 2007), vs. Epistemology computerized (Ganascia 2007). Discrete vs. continuous. Digital vs. Analog (Müller 2007). The classical Problem of Other Minds will help us to explore the systems in which computer "minds" approach the complexity of human minds. Cyborgs, human/machine combinations. Self-reflexive systems (Lanzarone 2007). Agent vs. environment. Bio-AI based on Bernie Baars' work.

- Philosophy of Computer Science

Pioneering contributions to Philosophy of Computer Science are courses done by Rapaport (2006), Tedre (2007), and research of Taylor and Eden (2007). Among others, the following philosophically significant questions will be addressed: What is a computer program? What is software?

- Philosophy of AI

Of all research fields of Computing, AI has the deepest connections to philosophy and with good reason. According to Chaitin (2007) you really understand something if you can program it, so we can say we really understand intelligence if we can program it. One could generalize "to program" into "to compute," meaning that computing may not be the same as programming, having in mind natural computation.

- Virtual-Real-Model-Simulation[3]

One of the characteristic features of computers and computing technologies are what Moor (1985) calls "logical malleability"—they are excellent for representation of information and in that capacity they may stand for both virtual worlds and the real one, and the distinction between the two might be difficult to tell. Ubiquitous computing is winning space and we understand that we more and more live in an infosphere (Floridi) which is radically different from the one of the era without ICT. That raises questions of the future use of computing in the production of virtual worlds and inspires investigations into the character of reality and the distinction real-virtual.

On the pragmatic side, there is a wide-spread use of computational models as a tool in natural and social sciences, humanities, engineering, government, etc. It is now well-recognized that we are witnessing a golden age of nanotechnologies with design of novel materials, and discoveries important for both basic science and applications. All mentioned are enabled by powerful computational tools.

We will discuss the role of computation and simulation in the dramatic advances of modeling and representation techniques we are witnessing today. Some specific present advances will be mentioned, such as quantum materials design, with the goal to synthesize in a controlled way materials on the atomic scale. The theory continues to develop along with computing power.

More examples of simulation will be found in robotics, artificial life, games, modeling of social systems, process monitoring software, and many others. Lanzarone (2007) presents an interesting view of Second Life (SL) as computational self-reflective system:

"The internal/external, observer/observed relationship is the basic concept of all virtual worlds.[11] In SL there seems to be a continuous interplay between in-world and out-world (jumping in and out of the system). In a certain sense, one could continuously enter and exit from the screen, or be at the same time on both sides of the screen. A sort of third life emerges from the interact ion between RL and SL."

- Ethics

This part of the course is divided in two.

1. Theoretical concepts
   - I. History of the term "computer ethics": Walter Manor, Jim Moor, Deborah Johnson
     Information Ethics Luciano Floridi (2007)
   - II. Philosophical meta-ethics and computing, professional ethics, micro/macro ethics: Don Gotterbarn
   - III. "Procedural ethics" for IT issues: just consequentialism, STS analysis, virtue ethics, and the importance of technical detail
2. Selected topics:
   - I. Privacy and IT
   - II. Killer robots in North Korea, Iraq, and downtown
   - III. Intellectual property
   - IV. Open source software: moral imperative?

- Computers in Society. Computers and Arts

The course will address the role of computers in society and arts, as a part of the answer to the questions what we can be done with and what can we expect from computing.

The relevance of this course for a computer science researcher or a student

The body of knowledge and practices in computing, as a new research field, has grown around an artifact—a computer. Unlike old research disciplines, especially physics, which has deep historical roots in Natural Philosophy, research tradition within the computing community up to now was primarily focused on problem solving and had not developed very strong bonds with philosophy.[4] The discovery of philosophical significance of computing in both philosophy and computing communities has led to a variety of new and interesting insights on both sides.

The view that information is the central idea of Computing/Informatics is both scientifically and sociologically indicative. Scientifically, it suggests a view of Informatics as a generalization of information theory that is concerned not only with the transmission/communication of information but also with its transformation and interpretation. Sociologically, it suggests a parallel between the industrial revolution, which is concerned with the utilizing of energy, and the information revolution, which is concerned with the utilizing of information (Dodig-Crnkovic 2003).

The relevance of this course for a philosophy researcher or a student

The development of philosophy is sometimes understood as its defining new research fields and then leaving them to sciences for further investigations (Floridi's lecture in Swedish National PI course on the development of philosophy, PI, 2004). At the same time, philosophy traditionally also learns from sciences and technologies, using them as tools for production of the

most reliable knowledge about the factual state of affairs of the world. We can mention a fresh example of current progress in modeling and simulation of brain and cognition that is of vital importance for the philosophy of mind. As so many times in history, the first best approach when scarce empirical knowledge exists, the intuitive one does not necessarily need to be the best. Wolpert (1993), for example, points out that science is an *unnatural mode of thought*, and it very often produces a counterintuitive knowledge, originating from the experiences with the world made by tools different from everyday ones, experiences in micro-cosmos, macro-cosmos, and other areas hidden for everyday experience. A good example of "unnatural" character of scientific knowledge is a totally counterintuitive finding of astronomy that earth is revolving around the sun. At present, similar Copernican Revolution seem to be going on in the philosophy of mind, epistemology (understood in informational terms), in philosophy of information, and philosophy of computing.

It is worth pointing out the novelty of the CaP course subject and scope. This is the first course of its kind, even though several courses have recently been developed internationally, addressing Computers and Philosophy, Philosophy of Computer Science (Rapaport 2006), and Philosophy of Information. At present we have established collaboration between several American and European universities with an ambition to develop an even wider global course in the future (CaP 2008).

Acknowledgements

The author would like to thank Peter Boltuc, Gaetano Lanzarone, Vincent Müller, and Keith Miller for reviewing the manuscript and offering valuable suggestions. Further credit and appreciation is extended to Bill Rapaport, Luciano Floridi, and Marvin Croy for their help regarding this project.

Endnotes

1. The majority of the above research (topics 1-7) will be addressed in the present course.
2. P. Thagard. "Societies of minds: Science as Distributed Computing. Studies in History and Philosophy of Science." 24 (1993): 49-67. Available at http://cogprints.org/676/0/Societies.html.
3. Excellent reading is (Lanzarone 2007) what gives a broader context of the relationship of real and virtual worlds as a relationship between inside and outside in an open interactive system.
4. Alan Turing was one of the notable exemptions to the rule. Others are Weizenbaum, Winograd, and Flores (GA Lanzarone, *APA Newsletter* Fall 2007). It should also be mentioned that Computing always had strong bonds with logic, and that especially AI always had recognized philosophical aspects.

References

Chaitin G.J. 2007. Epistemology as Information Theory. Alan Turing Lecture given at E-CAP 2005. In C*omputation, Information, Cognition – The Nexus and The Liminal*, edited by Dodig-Crnkovic G. and Stuart S. Cambridge Scholars Publishing, Cambridge UK. Available at http://www.cs.auckland.ac.nz/CDMTCS/chaitin/ecap.html.

Dodig-Crnkovic G. 2003. Shifting the Paradigm of the Philosophy of Science: The Philosophy of Information and a New Renaissance. *Minds and Machines* 13(4). Available at http://www.springerlink.com/content/g14t483510156726/ http://www.idt.mdh.se/personal/gdc/work/shifting_paradigm_singlespace.pdf.

Dodig-Crnkovic G. 2006. *Investigations into Information Semantics and Ethics of Computing.* Mälardalen University.

Available at http://www.idt.mdh.se/personal/gdc/work/publications.html.

Dodig-Crnkovic G. and Stuart S., eds. 2007. *Computation, Information, Cognition – The Nexus and The Liminal*. Cambridge Scholars Publishing, Cambridge UK. Available at http://www.idt.mdh.se/ECAP-2005/Intro-Preface-ComputationInformationCognition.pdf.

Dodig-Crnkovic G. 2007. Epistemology Naturalized: The Info-Computationalist Approach. *APA Newsletter on Philosophy and Computers* 06(2). Available at: http://www.apa.udel.edu/apa/publications/newsletters/v06n2/Computers/04.asp

Dodig-Crnkovic G. 2008. Semantics of Information as Interactive Computation, In *WSPI 2008, Kaiserslautern*, edited by Moeller, Neuser, and Roth-Berghofer. (DFKI Technical Reports; Berlin: Springer). Available at http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-332/paper4.pdf.

Dodig-Crnković G., Crnković I.. 2007. Increasing Interdisciplinarity by Distance Learning: Examples Connecting Economics with Software Engineering, and Computing with Philosophy. *E-mentor* 2(19).

www.e-mentor.edu.pl/eng CaP Computing and Philosophy global course http://www.idt.mdh.se/kurser/comphil/ (2008)

Floridi L. 2004. Open Problems in the Philosophy of Information. *Metaphilosophy* 35(4).

Floridi L. (ed), (2004) *The Blackwell Guide to the Philosophy of Computing and Information.* Blackwell Philosophy Guides.

Floridi L. 2007. Understanding Information Ethics. *APA Newsletter on Philosophy and Computers* 07(1). Available at http://www.apaonline.org/documents/publications/v07n1_Computers.pdf

Floridi L. 2005. Semantic Conceptions of Information. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Available at http://plato.stanford.edu/entries/information-semantic.

Floridi L. 2007. Information Logic. In *Computation, Information, Cognition. The Nexus and the Liminal*, edited by Dodig-Crnkovic G., Stuart S. Cambridge Scholars Press, Cambridge, UK.

Fredkin E. Digital Philosophy (references therein) http://www.digitalphilosophy.org/Home/Papers/tabid/61/Default.aspx.

Ganascia JG. 2008. "In silico" Experiments: Towards a Computerized Epistemology. *APA Newsletter on Philosophy and Computers* 07(2). Available at: http://www.apa.udel.edu/apa/publications/newsletters/v07n2/Computers/05.asp.

Goldin D, Smolka S, and Wegner P Ed. 2006. *Interactive Computation: The New Paradigm*. Springer-Verlag.

Halpin H. 2007. Philosophical Engineering: Towards a Philosophy of the Web. *APA Newsletter on Philosophy and Computers* 07(2):7-13. Available at http://www.apaonline.org/documents/publications/v07n2_Computers.pdf.

Lanzarone GA, Computing and Philosophy: In Search for a New Agenda. *APA Newsletter on Philosophy and Computers* 07(1):18-21.

Lloyd, S. 2006. *Programming the Universe: A Quantum Computer Scientist Takes on the Cosmos*. Jonathan Cape.

MacLennan B. 2004. Natural computation and non-Turing models of computation. *Theoretical Computer Science* 317:115-45.

Mainzer, K. 2004. *Thinking in complexity*. Springer, Berlin.

Mainzer, K. 2003. *Computerphilosophie zur Einführung*. Junius, Hamburg.

Milner R. 1989. *Communication and Concurrency*. Prentice-Hall.

Moor, J. 1985. What Is Computer Ethics? *Metaphilosophy* 16(4):266-75.

Müller V. 2007. Representation in digital systems. Interdisciplines: Adaptation and Representation. Available at http://www.interdisciplines.org/adaptation/papers.

PI. 2004. Swedish National Course, http://www.idt.mdh.se/~gdc/PI_04/index.html.

Priest, G., Tanaka, K. 2004. Paraconsistent Logic. *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Available at http://plato.stanford.edu/archives/win2004/entries/logic-paraconsistent/.

Rapaport, William J. 2005. Philosophy of Computer Science: An Introductory Course. *Teaching Philosophy* 28(4):319-41.

Siegelman, H.T. 1999. *Neural Networks and Analog Computation*. Birkhauser, Berlin.

Smith, J. and Jenks C. 2006. *Qualitative Complexity Ecology, Cognitive Processes and the Re-Emergence of Structures in Post-Humanist Social Theory*. Routledge.

Stonier T. 1993 *The Wealth of Information*. London: Thames/ Methuen.

Stonier T. 1997. *Information and Meaning. An Evolutionary Perspective*. Berlin, New York: Springer.

Tedre, M. 2007 Know Your Discipline: Teaching the Philosophy of Computer Science. *Journal of Information Technology Education* 6(1):105-22.

Turner R., Eden A.H. 2007. Towards a Programming Language Ontology, In *Computation, Information, Cognition—The Nexus and the Liminal*, edited by Gordana Dodig-Crnkovic, Susan Stuart. 147-59. Cambridge, UK: Cambridge Scholars Press.

Wegner P. 1998. Interactive Foundations of Computing. *Theoretical Computer Science* 192:315-51.

Wolfram S. 2002. *A New Kind of Science*. Wolfram Science.

Wolpert L. 1993. *The Unnatural Nature of Science*. Harvard University Press.

Zuse, K. 1967. *Rechnender Raum*. Elektronische Datenverarbeitung 8:336-44.

NOTES

Report on the International e-Learning Conference for Philosophy, Theology and Religious Studies, York, UK, May 14th-15th 2008

Constantinos Athanasopoulos
University of Leeds

The Subject Centre for Philosophical and Religious Studies, one of twenty-four subject centers located in higher education institutions throughout the UK, is part of the Higher Education Academy (http://www.heacademy.ac.uk/), and has a key role in the e-learning support and development of the university level philosophy teaching and learning in the UK. It recently organized with great success an International e-Learning Conference with participants from the UK, Italy, and Africa at York, UK, in May 2008. This two-day conference discussed ways to apply, embed, and enhance dialogue in applications of e-learning in Philosophy, Theology, Religious Studies, and cognate disciplines.

Dialogue has been frequently discussed as a major challenge for e-learning in the humanities (and especially in philosophy, HPS, theology, and religious studies). Roschelle (1992) identified the problem of *convergence* as one of the most important challenges in all collaborative attempts, especially the ones involving the use of ICT. More recently Alwood et al. (2000) claimed that recent development in the theory and applications of ICT has primarily helped in the *cognitive* considerations of embedding and enhancing human to human and human to computer dialogue and has done very little in supporting the other three key considerations in a successful human to human dialogue: joint purpose, ethics, and trust. Other projects, investigating how e-learning and dialogue can be brought together for the further enhancement of the educational process, have focused on the effects for learning from the study of computer mediated *vicarious learners* (cf. Lee et al. 1998; Craig et al. 2006). In addition, Walton (2000), while acknowledging the value of dialogue for the development of expert communication systems in both AI and teaching, claims that the information-seeking process in ICT expert systems can provide a distinctive framework of argumentation that can further enhance communication (and thus the educational process). Finally, Carusi (2003), while acknowledging that philosophers may have a xenophobic suspicion in taking a philosophical dialogue with their students online, claims that there are a lot of benefits in adopting ICT in the teaching of philosophy and especially in the way it can enhance dialogue through the text-based activities that it supports.

The conference provided a forum for discussing many aspects and issues related to the way that dialogue can be enhanced and supported via e-learning (especially in a deep-learning discipline such as philosophy). It included presentations and lectures from leading figures in e-learning in the humanities (from the UK, Italy, and Africa), who presented both traditional and pioneering approaches to e-learning.

Of special interest to philosophers were the following contributions:

a) Professor Luciano Floridi's discussion of the fourth revolution (as he called it) regarding the evolution of human culture led him to conclude that we have achieved a new status as information-based organisms or *inforgs*, who have very few differences from virtual agents or characters and with whom they share their infosphere. Floridi's challenging approach stimulated our thoughts and made us realize that innovations in the culture of ICT have produced an alternative viewpoint on our personhood that needs to be further examined and assessed. His insights are of special importance to the further e-learning development in our discipline, since more and more universities in UK and USA are moving into islands in Second Life and organize virtual campuses there, which are sometimes antagonistic to their real life programs and activities (Prof. Floridi's paper actually proved that the very distinction between the "real" and the "virtual" is not only outdated, but dangerously misleading…).

b) Dr. Catherine McCall presented her work on the theory and application of philosophical dialogue, comparing her own theory to Nelson's and Lipman's theories of dialogue.

c) Professor Richard Andrews presented his innovative theory about the relationship of dialogue and technology as applied in e-learning (presented in his book: *Sage Handbook of E-learning Research*).

d) Professor Dory Scaltsas presented the Archelogos Project (http://www.archelogos.com/), which he created and has run for more than twenty years now with great success, creating new content and semantic technologies, adapting and evolving the software to create both new content and new opportunities for further development and application of ICT in the analysis and assessment of philosophical argumentation, with a primary emphasis on the monumental works of Plato and Aristotle.

e) Roger Young proposed the creation of specific software for the enhancement of critical skills in the online teaching of an Introduction to Logic course.

f) Dr. Alex Zistakis discussed the theory of dialogue in the context of the platonic dialogues and how this can be applied to e-learning.

g) Dr. Annamaria Carusi discussed the value of the use of epistemological and semantic engines in the teaching of philosophy online and presented some of the challenges that teachers who use such engines may face.

h) Professor Livio Rosetti discussed his long experience with the creation of what he termed as "Metacognitive hypertexts," i.e., hypertexts that are based on metacognitive skills, an e-learning methodology which is based on his teaching of Plato's dialogues online or with DVD based software.

i) David Hunter presented his teaching experience using Lipman's theory of dialogue through the use of innovative tools such as blogs.

j) George Macdonald Ross presented his teaching experience in the use of MCQs as applied to a module on Kant's *Critique of Pure Reason* and the lessons that can be drawn from the incorporation of MCQs to support and enhance dialogue online.

k) Carl Smith presented the use of state-of-the-art innovative image and mobile communication technologies to create powerful online dialogues via images.

l) Mary Haight presented her methodology of teaching logic as a dialogue between characters and its potential for online use in the teaching introductory courses on logic.

One important aspect of the event was the ability of the participants to network and exchange ideas on further e-learning developments in philosophy. Key people from the world of e-learning in the UK (Dr. Lawrence Hamburg, Head of e-Learning in the Higher Education Academy) and Dr. Malcolm Read (JISC Executive Secretary) met and discussed with academics further e-learning developments in UK and about the possibilities of further funding in the area. It also provided the opportunity for additional information about the resources the SC, the Higher Education Academy, and JISC can provide to academics teaching philosophy in the UK. The conference presentations and archived webcasts from it are available at our conference webpage: http://prs.heacademy.ac.uk/projects/elearning/elearning_in_dialogue.html and our Wiki Resources webpage: http://wiki.prs.heacademy.ac.uk/doku.php?id=international_prs_e-learning_conference_in_york_14th-15th_may_2008, and there are plans of producing a volume of proceedings and a DVD with selected video recordings in the near future.

Bibliography

Allwood, Jens et al. "Cooperation, dialogue and ethics." *Int. J. Human-Computer Studies* 53 (2000): 871-914.

Carusi, Annamaria. "Project Report: Taking Philosophical Dialogue Online." *Discourse: Learning and Teaching in Philosophical and Religious Studies* 3:1 (2003): 95-156.

Craig Scotty D. et al. "The Deep-Level-Reasoning-Question Effect: The Role of Dialogue and Deep-Level-Reasoning Questions During Vicarious Learning." *Cognition and Instruction* 24:4 (2006): 565-91.

Lee John et al. "Supporting Student Discussions: It Isn't Just Talk." *Education and Information Technologies* 3 (1998): 217-29.

Roschelle, Jeremy. "Learning by Collaborating: Convergent Conceptual Change." *The Journal of the Learning Sciences* 2:3 (1992): 235-76.

Walton, Douglas. "The place of dialogue theory in logic, computer science and communication studies." *Synthese* 123 (2000): 327-46.

Call for Papers on The Ontological Status of Web-Based Objects

The *APA Newsletter on Philosophy and Computers* is seeking contributions on the topic of "The Ontological Status of Web-Based Objects." We hope for contributions taking up the issue from different, maybe new and unexpected angles; that's why we do not try to provide further directions.

Contributions, preferably of up to 3,000 words, should be emailed to the editor, to: pboltu@sgh.waw.pl.