BOOK REVIEW

Amy White: Virtually Obscene: The Case for an Uncensored Internet
Reviewed by Melissa Winkel

Amy White
"Reponse to Melissa Winkel"

NOTES

Cameron Buckner, Mathias Niepert, and Colin Allen
"InPhO: The Indiana Philosophy Ontology"

Geoffrey Klempner
"The Pathways School of Philosophy"

Matt Butcher
"NA-CAP@Loyola 2007"

APA NEWSLETTER ON

Philosophy and Computers

Piotr Bołtuć, Editor | Fall 2007 | Volume 07, Number 1

FROM THE EDITOR

Piotr Bołtuć
University of Illinois–Springfield

The previous issue of this Newsletter, the first one I edited, was devoted to making conspicuous the fact that philosophy and computers is no longer a new, controversial, or unproven field. We continue on along this path. In the current edition this point is made especially clearly in the article "Understanding Information Ethics" by Luciano Floridi, the current president of the International Association of Computing and Philosophy (IA-CAP), in which the author presents his main views developed since his 1999 book and spread among various articles.

It is hard to do justice to all aspects of this important paper in a few sentences; therefore, I will focus on the points I find of particular value for philosophy viewed broadly. The first point is the return of the ontological perspective in ethics (what Floridi calls re-ontologizing); this occurs through ontological, rather than just semantic or epistemological, interpretation of information. Information, including web-based objects, treated as equivalent to patterns or entities in the world, is a part of the metaphysical furniture of the world (though Floridi acknowledges that it belongs to a different level of abstraction than other objects). In fact, all entities, including humans, can be viewed as information objects. Human beings are in the process of rapid migration to the info-sphere and philosophy needs to acknowledge this fact more readily. This "ontological revolution" cannot but have far reaching implications in ethics.

Floridi presents ethics based upon something even more elemental than life, namely, being and upon something even more fundamental than suffering, namely, entropy. The latter has affinities with the concept of entropy used in thermodynamics but is also a bit like the metaphysical concept of nothingness. Moral theory becomes objectivist, and deeply non-anthropocentric, since all beings qua information objects have an intrinsic moral value. It becomes a moral imperative that entropy is to be diminished (and thus, anti-entropy enhanced). I am strongly inclined to believe that once this radical view reaches the radar screen of mainstream moral theorists the revolution it causes will become unstoppable. Floridi's view is a new paradigm in moral theory.

Riccardo Manzotti, in his article "Towards Artificial Consciousness," provides an overview of the field of artificial consciousness, which covers the issues between artificial intelligence (AI) and the traditional philosophy of mind. When presenting his own position, Manzotti follows up on his previous works in this area and argues for a process-oriented view of consciousness cast within an externalist framework of the mind. Manzotti's paper also serves as an excellent link to the second part of the Newsletter.

The next two papers comment on the keynote article from the last issue of this Newsletter, Gilbert Harman's "Explaining an Explanatory Gap." In his contribution, Yujin Nagasawa presents a critical analysis of the way Harman formulates the explanatory gap. He refers to Nagel's what it is like to be a bat example to show that the explanatory gap may emerge from one of two reasons: either we have to be bat-type creatures to know what it is like to have sonar, or objective characteristics of a bat do not tell us what it is like to have sonar. Nagasawa sketches out some implications of each alternative, which lead him to the conclusion that even if, per impossibile, there were a one-one correspondence between the phenomenal experience of a bat and that of a Homo sapiens, we would still not have a full physical characterization of what it is like to be a human being.

In her discussion of Harman's paper, Marion Ledwig claims that Harman gave no argument why the explanatory gap has no metaphysical implications; that it is not clear whether the explanatory gap is inevitable, or whether it could be bridged; that Harman's conception of Das Verstehen may need to include the social context. My favorite question posed by Ledwig is whether Harman's position allows "partial Verstehen."

Nagasawa and Ledwig gave a good start to our section of discussions and commentaries. I warmly encourage discussion of all papers published in this Newsletter and particularly of the featured articles (currently, Harman's and Floridi's).

G.A. Lanzarone's discussion paper, entitled "Computing and Philosophy: In Search for a New Agenda," examines two new issues that the author argues philosophers should take up. First, Lanzarone discusses "computational reflection" (following Feferman's reflection principle), which provides helpful insights into the old topic of first- and second-order logic. Lanzarone emphasizes the interplay of levels and meta-levels in AI. Next, he discusses a completely different area, the opportunities for philosophical analysis provided by Second Life, the system viewed as an interplay of the in-world and the out-world. The paper was presented at the NA-CAP conference in July 2007.

In his thorough analytical article Bertil Rolf, who presented at the 2007 E-CAP, discusses various criteria of testing educational benefits of computerized reasoning software; the work may also be of interest from the viewpoint of inductive theory. The paper focuses on intercontextual and intracontextual testing and the appropriate kinds of causal modeling. Intercontextual testing encounters difficulties which the author casts aptly in the example of functional comparison of axes versus saws. While in a few cases the relative effects of axes and saws can be compared, "most often, the user purpose and user process are not quite comparable." All we can conclude is that such and such tools "in the hands of such craftsmen, using such

techniques can bring forth" such and such outcomes. The same goes for educational testing, including the testing of educational software.

We return to the practice of publishing book reviews with that of Amy White's book Virtually Obscene: The Case for an Uncensored Internet by Melissa Winkel. We are also happy to publish the author's reply. The book is important since it contributes to the discussion of Internet freedom and provides strong arguments in its defense. White is a member of this committee; it is my intention to devote more space to the work of Committee members in future issues. The review is also important for the Newsletter to develop a broader section of book reviews; hence, uninvited submissions of book reviews, including those from graduate students, are warmly invited.

The Newsletter is glad to publish the review of various initiatives, programs, and websites devoted to philosophy and computers. In their note Colin Allen, Cameron Buckner, and Mathias Niepert give an updated presentation of InPhO: The Indiana Philosophy Ontology. The project consists of providing machine readable representation (ontology) of the relations among philosophical ideas. The project has theoretical implications since this level of formalization helps clarify philosophical statements, but it is geared primarily towards practical uses. It aims at providing conceptual navigation through the Stanford Encyclopedia of Philosophy online using not only semantic searches, but also information visualization techniques and other means. The project is part of a broader push towards digital philosophy.

Due to this editor's passion for online education in philosophy it is expected that each issue of the Newsletter will contain a presentation of one of the main educational projects online. We begin with an essay by Geoffrey Klempner, the editor of Pathways to Philosophy. The project has had a substantial web presence over the years and has encouraged many people to develop their philosophical skills outside of the rat race of academic degrees and accreditations.

Also, please find enclosed a note on this year's NA-CAP. NA-CAP is particularly important for the mission of this committee and we intend to keep our readers current on its work.

Last but not least, I need to mention that, right after this introduction, you shall find not just one, but two notes from the chair. This is because the current issue witnesses the transition of leadership in the APA Committee on Philosophy and Computers. Hence, the outgoing chair, Marvin Croy, focuses on the highlights of his term. Those include the evolution of the Barwise Prize into an important symbol of accomplishments in the field as well as gender equity among committee members. He also hopes that some form of a database documenting the uses of computers in philosophy, which the committee has been trying to establish for years on the APA website, shall eventually come to fruition.

The incoming chair, Michael Byron, highlights the research and the teaching strand of the committee and vouches to follow both strands in an equitable manner. Byron presents his past work in philosophy and computers, primarily focused on instructional software. He also gives a brief account of the two Committee sessions planned for the Eastern Division meeting in December 2007.

I would like to thank the Committee, and the APA, for providing me with the opportunity of editing the Newsletter and for making me the ex officio member of the Committee upon the expiration of my regular term this year; to Margot Duley, the Dean of Liberal Arts and Sciences at the University of Illinois (Springfield campus), for making it possible for me to devote some time and attention to this task; to John Barker, for friendly advice; and to my intern, Kaitlyn Patia, for her assistance. I want to give thanks to Ron Barnette, Keith Miller, and various philosophers all over the U.S. for serving as reviewers.

From the Outgoing Chair

Marvin Croy
University of North Carolina–Charlotte

This is my final report as chair of the PAC committee. In this capacity I succeeded Robert Cavalier (Carnegie Mellon University) and am being succeeded by Michael Byron (Kent State University). It seems like just yesterday that my term began. Actually, that was in July of 2003, and the years have disappeared amidst a busy and productive Committee schedule. During this interval the Committee sponsored over a dozen APA sessions, awarded the Barwise prize four times, produced numerous issues of the Newsletter, and strengthened its international ties, both by attracting international members and by continued collaboration with the International Association for Computing and Philosophy. I am very grateful for having been surrounded by and supported by a number of helpful resources. Primary among these have been the Committee members themselves. During my term, Committee members have included the following (generally in order of service):

S.D. Noam Cook (San Jose State University)
James H. Fetzer (University of Minnesota–Duluth)
Luciano L. Floridi (Oxford University)
Patrick N. Grim (State University of New York–Stony Brook)
Mark Manion (Drexel University)
David G. Stern (University of Iowa)
Bruce Umbaugh (Webster University)
Jon Dorbolo (Oregon State University)
Peter Boltuć (University of Illinois–Springfield)
Christopher Grau (Clemson University)
Branden Fitelson (University of California–Berkeley)
Susan Stuart (University of Glasgow)
Ange Cooksey (Indiana University–East)
Jerry Kapus (University of Wisconsin–Stout)
Amy White (Ohio University)
Harriet Baber (University of San Deigo)
Michael Byron (Kent State University; associate chair)

I am also extremely grateful for the assistance given by the APA staff in the National Office and the officers of the APA Divisions. Their assistance, encouragement, and friendship have sustained my work and that of the Committee, and has meant much to me personally. I am confident that their competence and generosity will be a substantial benefit to Michael Byron as he takes over the helm of the Committee. Another change involves the production of the Committee's Newsletter. Peter Boltuć has been serving as co-editor and now will assume the role of Newsletter editor. My thanks to Ange Cooksey for her service in editing the Newsletter over the last couple of years. Given the leadership of Michael and Peter, I am confident that the Committee will successfully take up new challenges in the years to come.

As with any endeavor, my service as chair has provided both satisfaction and frustration. One Committee accomplishment that I am proud of involves the evolution of the Barwise Prize.

Recipients include Pat Suppes, Dan Dennett, Deborah Johnson, Hubert Dreyfus, and Jim Moor. This prize has become an emblem that defines the Committee and highlights its mission. I am also proud of the changing gender composition of the Committee. The Committee has previously included women members, but when I assumed the chair, there were none. At this writing, women compose approximately 40 percent of the Committee, and given recent nominations, I expect that number to increase.

The Committee's central charge is to generate and communicate information concerning the uses of computers within philosophy. Currently this objective is primarily achieved via sessions at APA meetings, its Newsletter articles, and a few APA webpages. During my first year as chair, a design for a more elaborate system was produced. This system for collecting and reporting relevant information was modeled on the APA's Grad Guide, which resides on the APA web server. Other committees, particularly the Committee on Teaching, expressed an interest in making use of the reports to be issued by this system. For several reasons, this system never materialized. I trust that it, or some functional equivalent, will become a reality.

Once again, I thank the Committee members for their contributions and efforts. I shall remember my years as PAC committee chair with warmth, and I wish the Committee success in all of its endeavors.

From the Incoming Chair

Michael Byron
Kent State University

I would like to begin by thanking Marvin Croy for his invitation to chair the committee, for his hard work as chair over the past three years, and for the advice he has given and will give to me. I'm sure that I speak for the Committee in expressing my gratitude for his contribution.

As I see the APA Committee on Philosophy and Computers, the group has two main strands. We might call these the research strand and the teaching strand, and these overlap at a growing number of points. The research strand comprises folks whose areas of specialization include anything to do with computers: philosophy of mind, AI, and artificial life would be paradigm instances. The teaching strand includes those of us interested in instructional software, distance learning, and technology in the service of teaching, broadly construed. Nothing grand hangs on this distinction, which in any case is hardly sharp. I mention it in order to notice that sometimes divergent interests bring people to our committee, and that the Committee's agenda should speak to as many of those interests as possible.

For my part, instructional software brought me to the committee. Back in 1999 I began a project to develop a distance learning version of Kent State's formal logic class, and I reviewed the main software packages available for that course. I later presented this work at the CAP conference at Carnegie Mellon and published the (now quite dated) results in Teaching Philosophy. I am still teaching formal logic using instructional software, though no longer via distance learning. To further establish my computing bona fides, I have been a technophile and computer owner since the 1980s. My first computer was an Epson 128K machine that I bought when I started grad school in 1988. More recently: I designed and built the websites of the Ohio Philosophical Association (http://ohiophilosophy.org) and the Kent State philosophy department (http://philosophy.kent.edu), using HTML, PHP, MySQL, and any number of other letters. In short, I am a nerd.

In the coming year, the committee expects to host sessions at all three Divisional meetings. We have two sessions on the table for the Eastern Division meeting, December 27-30, 2007, in Baltimore. The first, proposed by Harriet Baber of the University of San Diego, is entitled "Technology in Support of Philosophy Research: Tools, Semantics, and Ontology." This panel discussion will include Robert Rynasiewicz and Sayeed Chaudhury of Johns Hopkins, and Bill Anderson of Ontology Works, Inc.

The second session at the Eastern Division meeting, proposed by Marvin Croy of the University of North Carolina–Charlotte, is entitled "The Ethics of Emerging Technologies." Speakers will include Marvin, Harriet, and Andrew Light, who is from the University of Washington. The topics to be addressed in this session concern the ethics of intelligent tutoring systems, access to information, and sustainability.

I look forward to working with the Committee in the coming years. I would like to close by thanking Peter Boltuc for agreeing to edit the Newsletter. Judging by his efforts to get this note from me, he'll do a terrific job. Just remember, Peter: if it weren't for deadlines, a lot of work would never get done!

Articles

Understanding Information Ethics

Luciano Floridi
University of Oxford

1. Introduction

The informational revolution has been changing the world profoundly and irreversibly for more than half a century now, at a breathtaking pace and with an unprecedented scope. In a recent study on the evolution of information,[1] researchers at Berkeley's School of Information Management and Systems estimated that humanity had accumulated approximately 12 exabytes of data in the course of its history, but that the world had produced more than 5 exabytes of data just in 2002. This is almost 800 MB of recorded information produced per person each year. It is like saying that every new born baby comes to the world with a burden of 30 feet of books, the equivalent of 800 MB of information on paper. Most of these data are of course digital: 92 percent of them were stored on magnetic media, mostly in individuals' hard disks (the phenomenon is known as the "democratization" of data). So, hundreds of millions of computing machines are constantly employed to cope with exabytes of data. In 2005, they were more than 900M. By the end of 2007, it is estimated that there will be over 1.15B PCs in use, at a compound annual growth of 11.4 percent.[2] Of course, PCs are among the greatest sources of further exabytes.

All these numbers will keep growing for the foreseeable future. The result is that information and communication technologies (ICTs) are building the new informational habitat (what I shall define below as the infosphere) in which future generations will spend most of their time. In 2007, for example, it is estimated that American adults and teens will spend on average 3,518 waking hours inside the infosphere, watching television, surfing the Internet, reading daily newspapers, and

listening to personal music devices.[3] This is a total amount of nearly five months. Most of the remaining seven months will be spent eating, sleeping, using cell phones or other communication devices, and playing video games (already 69 percent of American heads of households play computer and video games).[4]

Building a worldwide, ethical infosphere, a fair digital habitat for all, raises unprecedented challenges for humanity in the twenty-first century. The U.S. Department of Commerce and the U.S. National Science Foundation have identified "NBIC" (Nanotechnology, Biotechnology, Information Technology, and Cognitive Science) as a national priority area of research and have recently sponsored a report entitled "Converging Technologies for Improving Human Performance." And in March 2000, the EU Heads of States and Governments acknowledged the radical transformations brought about by ICT when they agreed to make the EU "the most competitive and dynamic knowledge-driven economy by 2010."

Information and Communication Technologies and the information society are bringing concrete and imminent opportunities for enormous benefit to people's education, welfare, prosperity, and edification, as well as great economic advantages. But they also carry significant risks and generate moral dilemma and profound philosophical questions about human nature, the organization of a fair society, the "morally good life," and our responsibilities and obligations to present and future generations. In short, because the informational revolution is causing an exponential growth in human powers to understand, shape, and control ever more aspects of reality, it is equally making us increasingly responsible, morally speaking, for the way the world is, will, and should be, and for the role we are playing as stewards of our future digital environment. The informationalization of the world, of human society, and of ordinary life has created entirely new realities, made possible unprecedented phenomena and experiences, provided a wealth of extremely powerful tools and methodologies, raised a wide range of unique problems and conceptual issues, and opened up endless possibilities hitherto unimaginable. As a result, it has also deeply affected our moral choices and actions, affected the way in which we understand and evaluate moral issues, and posed fundamental ethical problems, whose complexity and dimensions are rapidly growing and evolving. It would not be an exaggeration to say that many ethical issues are related to or dependent on the computer revolution.

In this paper, I will look at the roots of the problem: what sort of impact ICTs are having or will soon have on our lives, and what kind of new ethical scenarios such technological transformations are ushering in. For this purpose, it will be convenient to explain immediately two key concepts and then outline the main claim that will be substantiated and explained in the following pages.

The first concept is that of infosphere, a neologism I coined in the nineties[5] on the basis of "biosphere," a term referring to that limited region on our planet that supports life. "Infosphere" denotes the whole informational environment constituted by all informational entities (thus also including informational agents like us or like companies, governments, etc.), their properties, interactions, processes, and mutual relations. It is an environment comparable to but different from cyberspace (which is only one of its sub-regions, as it were), since it also includes offline and analogue spaces of information. We shall see that it is also an environment (and hence a concept) that is rapidly evolving. The alerted reader will notice a (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).

The second concept is that of re-ontologization, another neologism that I have recently introduced in order to refer to a very radical form of re-engineering, one that not only designs, constructs, or structures a system (e.g., a company, a machine, or some artefact) anew, but that fundamentally transforms its intrinsic nature. In this sense, for example, nanotechnologies and biotechnologies 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.

Using the two previous concepts, my basic claims can now be formulated thus: computers and, more generally, digital ICTs are re-ontologizing the very nature of (and hence what we mean by) the infosphere; here lies the source of some profound ethical transformations and challenging problems; and Information Ethics (IE), understood as the philosophical foundation of Computer Ethics, can deal successfully with such challenges.

Unpacking these claims will require two steps. In sections 2-4, I will first analyze three fundamental trends in the re-ontologization of the infosphere.[6] This step should provide a sufficiently detailed background against which the reader will be able to evaluate the nature and scope of Information Ethics. In section 5, I will then introduce Information Ethics itself. I say "introduce" because the hard and detailed work of marshalling arguments and replies to objections will have to be left to the specialized literature.[7] Metaphorically, the goal will be to provide a taste of Information Ethics, not the actual recipes. Some concluding remarks in section 6 will close this paper.

2. The rise of the frictionless infosphere

The most obvious way in which the new ICTs are re-ontologizing the infosphere concerns the transition from analogue to digital data and then the ever-increasing growth of our digital space. This radical re-ontologization of the infosphere is largely due to the fundamental convergence between digital resources and digital tools. The ontology of the ICTs available (e.g., software, databases, communication channels and protocols, etc.) is now the same as (and hence fully compatible with) the ontology of their objects. This was one of Turing's most consequential intuitions: in the re-ontologized infosphere, there is no longer any substantial difference between the processor and the processed, so the digital deals effortlessly and seamlessly with the digital. This potentially eliminates one of the most long-standing bottlenecks in the infosphere and, as a result, there is a gradual erasure of ontological friction.

Ontological friction refers to the forces that oppose the flow of information within (a region of) the infosphere and, hence, (as a coefficient) to the amount of work and effort required to generate, obtain, process, and transmit information in a given environment, e.g., by establishing and maintaining channels of communication and by overcoming obstacles in the flow of information such as distance, noise, lack of resources (especially time and memory), amount and complexity of the data to be processed, and so forth. Given a certain amount of information available in (a region of) the infosphere, the lower the ontological friction in it, the higher the accessibility of that amount of information becomes. Thus, if one could quantify ontological friction from 0 to 1, a fully successful firewall would produce a 1.0 degree of friction for any unwanted connection, i.e., a complete standstill in the flow of the unwanted data through its "barrier." On the other hand, we describe our society as informationally porous the more it tends towards a 0 degree of informational friction.

Because of their "data superconductivity," ICTs are well known for being among the most influential factors that affect the

ontological friction in the infosphere. We are all acquainted with daily aspects of a frictionless infosphere, such as spamming and micropayments. Three other significant consequences are:

a) no right to ignore: in an increasingly porous society, it will become progressively less credible to claim ignorance when confronted by easily predictable events (e.g., as George W. Bush did with respect to Hurricane Katrina's disastrous effects on New Orleans's flood barriers) and painfully obvious facts (e.g., as British politician Tessa Jowell did with respect to her husband's finances in a widely publicized scandal)[8]; and

b) vast common knowledge: this is a technical term from epistemic logic, which basically refers to the case in which everybody not only knows that p but also knows that everybody knows that everybody knows that p. In other words, (a) will also be the case because meta-information about how much information is, was, or should have been available will become overabundant.

From (a) and (b) it follows that, in the future,

c) we shall witness a steady increase in agents' responsibilities. In particular, ICTs are making human agents increasingly accountable, morally speaking, for the way the world is, will, and should be.[9]

3. The global infosphere or how information is becoming our ecosystem

During the last decade or so, we have become accustomed to conceptualizing our life online as a mixture between an evolutionary adaptation of human agents to a digital environment and a form of post-modern, neo-colonization of the latter by the former. This is probably a mistake. Computers are as much re-ontologizing our world as they are creating new realities. The threshold between here (analogue, carbon-based, off-line) and there (digital, silicon-based, online) is fast becoming blurred, but this is as much to the advantage of the latter as it is of the former. This recent phenomenon is variously known as "Ubiquitous Computing," "Ambient Intelligence," "The Internet of Things" (ITU report, November 2005, http://www.itu.int/internetofthings) or "Web-augmented things." It is or will soon be the next stage in the digital revolution. To put it dramatically, the infosphere is progressively absorbing any other space. Let me explain.

In the (fast approaching) future, more and more objects will be what I'd like to call ITentities, able to learn, advise, and communicate with each other. A good example is provided by Radio Frequency Identification (RFID) tags, which can store and remotely retrieve data from an object and give it a unique identity, like a barcode. Tags can measure less than half a millimeter square and are thinner than paper. Incorporate this tiny microchip in everything, including humans and animals, and you have created ITentities. This is not science fiction. According to a report by Market Research Company InStat, the worldwide production of RFID will increase more than 25-fold between 2005 and 2010 and reach 33 billion. Imagine networking these 33 billion ITentities together with all the hundreds of millions of PCs, DVDs, iPods, and ICT devices available and you see that the infosphere is no longer "there" but "here" and it is here to stay. Your Nike and iPod already talk to each other (http://www.apple.com/ipod/nike/).

Nowadays, we are used to considering the space of information as something we log-in to and log-out from. Our view of the world (our metaphysics) is still modern or Newtonian: it is made of "dead" cars, buildings, furniture, clothes, which are non-interactive, irresponsive, and incapable of communicating, learning, or memorizing. But what we still experience as the world offline is bound to become a fully interactive and responsive environment of wireless, pervasive, distributed, a2a (anything to anything) information processes, that works a4a (anywhere for anytime), in real time. This will first gently invite us to understand the world as something "a-live" (artificially live). Such animation of the world will, paradoxically, make our outlook closer to that of pre-technological cultures, which interpreted all aspects of nature as inhabited by teleological forces.

The second step will be a reconceptualization of our ontology in informational terms. It will become normal to consider the world as part of the infosphere, not so much in the dystopian sense expressed by a Matrix-like scenario, where the "real reality" is still as hard as the metal of the machines that inhabit it; but in the evolutionary, hybrid sense represented by an environment such as New Port City, the fictional, post-cybernetic metropolis of Ghost in the Shell (http://en.wikipedia.org/wiki/Ghost_in_the_Shell). This is the shift I alerted you to some pages ago. The infosphere will not be a virtual environment supported by a genuinely "material" world behind; rather, it will be the world itself that will be increasingly interpreted and understood informationally, as part of the infosphere. At the end of this shift, the infosphere will have moved from being a way to refer to the space of information to being synonymous with Being or reality. This is the sort of informational metaphysics I suspect we shall find increasingly easy to embrace. Just ask one of the more than 8 million players of War of Warcraft, one of the almost 7 million inhabitants of Second Life, or one of the 70 million owners of Neopets.

4. The evolution of inforgs

We have seen that we are probably the last generation to experience a clear difference between "onlife" and online. The third transformation that I wish to highlight concerns precisely the emergence of artificial and hybrid (multi)agents, i.e., partly artificial and partly human. Consider, for example, a whole family as a single agent, equipped with digital cameras, laptops, Palm OS handhelds, iPods, mobile phones, camcorders, wireless networks, digital TVs, DVDs, CD players, and so on.

These new agents already share the same ontology with their environment and can operate in it with much more freedom and control. We (shall) delegate or outsource to artificial agents memories, decisions, routine tasks, and other activities in ways that will be increasingly integrated with us and with our understanding of what it means to be an agent. This is rather well known, but two other aspects of this transformation may be in need of some clarification.

On the one hand, in the re-ontologized infosphere, progressively populated by ontologically equal agents, where there is no difference between processors and processed, online and offline, all interactions become equally digital. They are all interpretable as "read/write" (i.e., access/alter) activities, with "execute" the remaining type of process. It is easy to predict that, in such an environment, the moral status and accountability of artificial agents will become an ever more challenging issue (Floridi and Sanders 2004b).

On the other hand, our understanding of ourselves as agents will also be deeply affected. I am not referring here to the sci-fi vision of a "cyborged" humanity. Walking around with something like a Bluetooth wireless headset implanted in your ear does not seem the best way forward, not least because it contradicts the social message it is also meant to be sending: being on call 24×7 is a form of slavery, and anyone so busy and important should have a PA instead. The truth is rather that being a sort of cyborg is not what people will embrace, but what they will try to avoid, unless it is inevitable (more on this shortly).

Nor am I referring to a genetically modified humanity, in charge of its informational DNA and, hence, of its future

embodiments. This is something that we shall probably see in the future, but it is still too far away, both technically (safely doable) and ethically (in the sense of being morally acceptable as normally as having a heart by-pass or some new spectacles: we are still struggling with the ethics of stem cells), to be discussed at this stage.

What I have in mind is a quieter, less sensational, and yet crucial and profound change in our conception of what it means to be an agent. We are all becoming connected informational organisms (inforgs). This is happening not through some fanciful transformation in our body but, more seriously and realistically, through the re-ontologization of our environment and of ourselves.

By re-ontologizing the infosphere, digital ICTs have brought to light the intrinsically informational nature of human agents. This is not equivalent to saying that people have digital alter egos, some Messrs Hydes represented by their @s, blogs, and https. This trivial point only encourages us to mistake digital ICTs for merely enhancing technologies. The informational nature of agents should not be confused with a "data shadow"[10] either. The more radical change, brought about by the re-ontologization of the infosphere, will be the disclosure of human agents as interconnected, informational organisms among other informational organisms and agents.

Consider the distinction between enhancing and augmenting appliances. The switches and dials of the former are interfaces meant to plug the appliance in to the user's body ergonomically. Drills and guns are perfect examples. It is the cyborg idea. The data and control panels of augmenting appliances are instead interfaces between different possible worlds: on the one hand there is the human user's Umwelt,[11] Euclidean, Newtonian, colorful, and so forth, and on the other hand there is the dynamic, watery, soapy, hot, and dark world of the dishwasher; the equally watery, soapy, hot, and dark but also spinning world of the washing machine; or the still, aseptic, soapless, cold, and potentially luminous world of the refrigerator. These robots can be successful because they have their environments "wrapped" and tailored around their capacities, not vice versa. Imagine someone trying to build a droid like C3PO capable of washing their dishes in the sink exactly in the same way as a human agent would.

Now, ICTs are not augmenting or empowering in the sense just explained. They are re-ontologizing devices because they engineer environments that the user is then enabled to enter through (possibly friendly) gateways. It is a form of initiation. Looking at the history of the mouse (http://sloan.stanford.edu/mousesite/), for example, one discovers that our technology has not only adapted to, but also educated, us as users. Douglas Engelbart once told me that he had even experimented with a mouse to be placed under the desk, to be operated with one's leg, in order to leave the user's hands free. Human-Computer Interaction (HCI) is a symmetric relation of mutual symbiosis.

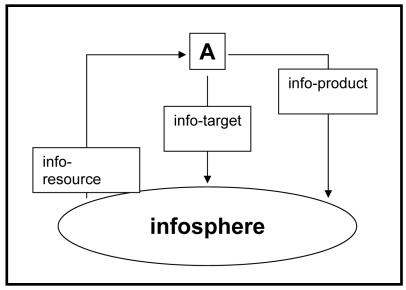
To return to our distinction, whilst a dishwasher interface is a panel through which the machine enters into the user's world, a digital interface is a gate through which a user can be (tele)present in the infosphere (Floridi 2005b). This simple but fundamental difference underlies the many spatial metaphors of "cyberspace," "virtual reality," "being online," "surfing the web," "gateway," and so forth. It follows that we are witnessing an epochal, unprecedented migration of humanity from its Umwelt to the infosphere itself, not least because the latter is absorbing the former. As a result, humans will be inforgs among other (possibly artificial) inforgs and agents operating in an environment that is friendlier to digital creatures. As digital immigrants like us are replaced by digital natives like our children, the latter will come to appreciate that there is no ontological difference between infosphere and Umwelt, only a difference of levels of abstractions (Floridi and Sanders 2004a). And when the migration is complete, we shall increasingly feel deprived, excluded, handicapped, or poor to the point of paralysis and psychological trauma whenever we are disconnected from the infosphere, like fish out of water.

5. Information Ethics as a new environmental ethics

In the previous sections, we have seen some crucial transformations brought about by ICT in our lives. Moral life is a highly information-intensive activity, so any technology that radically modifies the "life of information" is bound to have profound moral implications for any moral agent. Recall that we are talking about an ontological revolution, not just a change in communication technologies. ICTs, by radically transforming the informational context in which moral issues arise, not only add interesting new dimensions to old problems, but lead us to rethink, methodologically, the very grounds on which our ethical positions are based.[12] Let us see how.

ICTs affect an agent's moral life in many ways. For the sake of simplicity, they can be schematically organized along three lines (see Figure 1), in the following way.

Figure 1. The "External" R(esource) P(roduct) T(arget) Model



Suppose our moral agent A is interested in pursuing whatever she considers her best course of action, given her predicament. We shall assume that A's evaluations and interactions have some moral value, but no specific value needs to be introduced at this stage. Intuitively, A can avail herself of some information (information as a resource) to generate some other information (information as a product) and, in so doing, affect her informational environment (information as target). This simple model, summarized in Figure 1, may help one to get some initial orientation in the multiplicity of issues belonging to Information Ethics.[13] I shall refer to it as the RPT model.

The RPT model is useful to rectify an excessive emphasis occasionally placed on specific technologies (this happens most notably in computer ethics) by calling our attention to the more fundamental phenomenon of information in all its varieties and long tradition. This was also Wiener's position[14] and the various difficulties encountered in the conceptual foundations of computer ethics are arguably[15] connected to the fact that the latter has not yet been recognized as primarily an environmental ethics, whose main concern is (or should be) the ecological management and well being of the infosphere.

Since the appearance of the first works in the eighties,[16] Information Ethics has been claimed to be the study of moral issues arising from one or another of the three distinct "information arrows" in the RPT model. This is not entirely satisfactory.

5.1. Information-as-a-resource Ethics

Consider first the crucial role played by information as a resource for A's moral evaluations and actions. Moral evaluations and actions have an epistemic component, since A may be expected to proceed "to the best of her information," that is, A may be expected to avail herself of whatever information she can muster, in order to reach (better) conclusions about what can

and ought to be done in some given circumstances. Socrates already argued that a moral agent is naturally interested in gaining as much valuable information as the circumstances require, and that a well-informed agent is more likely to do the right thing. The ensuing "ethical intellectualism" analyzes evil and morally wrong behavior as the outcome of deficient information. Conversely, A's moral responsibility tends to be directly proportional to A's degree of information: any decrease in the latter usually corresponds to a decrease in the former. This is the sense in which information occurs in the guise of judicial evidence. It is also the sense in which one speaks of A's informed decision, informed consent, or well-informed participation. In Christian ethics, even the worst sins can be forgiven in the light of the sinner's insufficient information, as a counterfactual evaluation is possible: had A been properly informed A would have acted differently and hence would not have sinned (Luke 23:34). In a secular context, Oedipus and Macbeth remind us how the mismanagement of informational resources may have tragic consequences.[17]

From a "resource" perspective, it seems that the moral machine needs information, and quite a lot of it, to function properly. However, even within the limited scope adopted by an analysis based solely on information as a resource and, hence, a merely semantic view of the infosphere, care should be exercised, lest all ethical discourse is reduced to the nuances of higher quantity, quality, and intelligibility of informational resources. The more the better is not the only, nor always the best, rule of thumb. For the (sometimes explicit and conscious) withdrawal of information can often make a significant difference. A may need to lack (or preclude herself from accessing) some information in order to achieve morally desirable goals, such as protecting anonymity, enhancing fair treatment, or implementing unbiased evaluation. Famously, Rawls' "veil of ignorance" exploits precisely this aspect of information-as-a-resource, in order to develop an impartial approach to justice (Rawls 1999). Being informed is not always a blessing and might even be morally wrong or dangerous.

Whether the (quantitative and qualitative) presence or the (total) absence of information-as-a-resource is in question, it is obvious that there is a perfectly reasonable sense in which Information Ethics may be described as the study of the moral issues arising from "the triple A": availability, accessibility, and accuracy of informational resources, independently of their format, kind, and physical support.[18] Rawls' position has been already mentioned. Other examples of issues in IE, understood as an Information-as-resource Ethics, are the so-called digital divide, the problem of infoglut, and the analysis of the reliability and trustworthiness of information sources.

5.2. Information-as-a-product Ethics

A second but closely related sense in which information plays an important moral role is as a product of A's moral evaluations and actions. A is not only an information consumer but also an information producer, who may be subject to constraints while being able to take advantage of opportunities. Both constraints and opportunities call for an ethical analysis. Thus, IE, understood as Information-as-a-product Ethics, may cover moral issues arising, for example, in the context of accountability, liability, libel legislation, testimony, plagiarism, advertising, propaganda, misinformation, and more generally of pragmatic rules of communication à la Grice. Kant's analysis of the immorality of lying is one of the best known case studies in the philosophical literature concerning this kind of Information Ethics. Cassandra and Laocoon, pointlessly warning the Trojans against the Greeks' wooden horse, remind us how the ineffective management of informational products may have tragic consequences.

5.3. Information-as-a-target Ethics

Independently of A's information input (info-resource) and output (info-product), there is a third sense in which information may be subject to ethical analysis, namely, when A's moral evaluations and actions affect the informational environment. Think, for example, of A's respect for, or breach of, someone's information privacy or confidentiality.[19] Hacking, understood as the unauthorized access to a (usually computerized) information system, is another good example. It is not uncommon to mistake it for a problem to be discussed within the conceptual frame of an ethics of informational resources. This misclassification allows the hacker to defend his position by arguing that no use (let alone misuse) of the accessed information has been made. Yet hacking, properly understood, is a form of breach of privacy. What is in question is not what A does with the information, which has been accessed without authorization, but what it means for an informational environment to be accessed by A without authorization. So the analysis of hacking belongs to an Info-target Ethics. Other issues here include security, vandalism (from the burning of libraries and books to the dissemination of viruses), piracy, intellectual property, open source, freedom of expression, censorship, filtering, and contents control. Mill's analysis "Of the Liberty of Thought and Discussion" is a classic of IE interpreted as Information-as-target Ethics. Juliet, simulating her death, and Hamlet, re-enacting his father's homicide, show how the risky management of one's informational environment may have tragic consequences.

5.4. The limits of any microethical approach to Information Ethics

At the end of this overview, it seems that the RPT model may help one to get some initial orientation in the multiplicity of issues belonging to different interpretations of Information Ethics. Despite its advantages, however, the model can still be criticized for being inadequate in two respects.

On the one hand, the model is too simplistic. Arguably, several important issues belong mainly but not only to the analysis of just one "informational arrow." The reader may have already thought of several examples that illustrate the problem: someone's testimony is someone's else trustworthy information; A's responsibility may be determined by the information A holds, but it may also concern the information A issues; censorship affects A both as a user and as a producer of information; misinformation (i.e., the deliberate production and distribution of false and misleading contents) is an ethical problem that concerns all three "informational arrows"; freedom of speech also affects the availability of offensive content (e.g., child pornography, violent content, and socially, politically, or religiously disrespectful statements) that might be morally questionable and should not circulate.

On the other hand, the model is insufficiently inclusive. There are many important issues that cannot easily be placed on the map at all, for they really emerge from, or supervene on, the interactions among the "informational arrows." Two significant examples may suffice: "big brother," that is, the problem of monitoring and controlling anything that might concern A; the debate about information ownership (including copyright and patents legislation) and fair use, which affects both users and producers while shaping their informational environment.

So the criticism is reasonable. The RPT model is indeed inadequate. Yet why it is inadequate is a different matter. The tripartite analysis just provided is unsatisfactory, despite its partial usefulness, precisely because any interpretation of Information Ethics based on only one of the "informational arrows" is bound to be too reductive. As the examples mentioned above emphasize, supporters of narrowly constructed interpretations

of Information Ethics as a microethics (that is, a practical, field-dependent, applied, and professional ethics) are faced by the problem of being unable to cope with a large variety of relevant issues (I mentioned some of them above), which remain either uncovered or inexplicable. In other words, the model shows that idiosyncratic versions of IE, which privilege only some limited aspects of the information cycle, are unsatisfactory. We should not use the model to attempt to pigeonhole problems neatly, which is impossible. We should rather exploit it as a useful scheme to be superseded, in view of a more encompassing approach to IE as a macroethics, that is, a theoretical, field-independent, applicable ethics. Philosophers will recognize here a Wittgensteinian ladder, which can be used to reach a new starting point, but then can be discharged.

In order to climb up on, and then throw away, any narrowly constructed conception of Information Ethics, a more encompassing approach to IE needs to

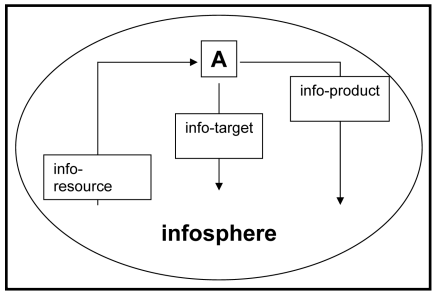
i) bring together the three "informational arrows";

ii) consider the whole information-cycle; and

iii) take seriously the ontological shift in the nature of the infosphere that I emphasized above, thus analyzing informationally all entities involved (including the moral agent A) and their changes, actions, and interactions, treating them not apart from, but as part of the informational environment to which they belong as informational systems themselves.

Whereas steps (i) and (ii) do not pose particular problems and may be shared by other approaches to IE, step (iii) is crucial but involves an "update" in the ontological conception of "information" at stake. Instead of limiting the analysis to (veridical) semantic contents—as any narrower interpretation of IE as a microethics inevitably does—an ecological approach to Information Ethics also looks at information from an object-oriented perspective and treats it as an entity as well. In other words, we move from a (broadly constructed) epistemological or semantic conception of Information Ethics—in which information is roughly equivalent to news or contents—to one which is typically ontological, and treat information as equivalent to patterns or entities in the world. Thus, in the revised RPT model, represented in Figure 2, the agent is embodied and embedded, as an informational agent, in an equally informational environment.

Figure 2. "Internal" R(esource) P(roduct) T(arget) Model: the Agent A is correctly embedded within the infosphere.



A simple analogy may help to introduce this new perspective.[20] Imagine looking at the whole universe from a chemical perspective.[21] Every entity and process will satisfy a certain chemical description. A human being, for example, will be 70 percent water and 30 percent something else. Now consider an informational perspective. The same entities will be described as clusters of data, that is, as informational objects. More precisely, our agent A (like any other entity) will be a discrete, self-contained, encapsulated package containing

i) the appropriate data structures, which constitute the nature of the entity in question, that is, the state of the object, its unique identity, and its attributes; and

ii) a collection of operations, functions, or procedures, which are activated by various interactions or stimuli (that is, messages received from other objects or changes within itself) and correspondingly define how the object behaves or reacts to them.

At this level of analysis, informational systems as such, rather than just living systems in general, are raised to the role of agents and patients of any action, with environmental processes, changes, and interactions equally described informationally.

Understanding the nature of IE ontologically rather than epistemologically modifies the interpretation of the scope of IE. Not only can an ecological IE gain a global view of the whole life-cycle of information, thus overcoming the limits of other microethical approaches, but it can also claim a role as a macroethics, that is, as an ethics that concerns the whole realm of reality. This is what we shall see in the next section.

5.5. Information Ethics as a Macroethics

Information Ethics is patient-oriented, ontocentric, ecological macroethics (Floridi 1999a; Floridi and Sanders 1999). These are technical expressions that can be intuitively explained by comparing IE to other environmental approaches.

Biocentric ethics usually grounds its analysis of the moral standing of bio-entities and eco-systems on the intrinsic worthiness of life and the intrinsically negative value of suffering. It seeks to develop a patient-oriented ethics in which the "patient" may be not only a human being, but also any form of life. Indeed, Land Ethics extends the concept of patient to any component of the environment, thus coming close to the approach defended by Information Ethics. Rowlands (2000), for example, has recently proposed an interesting approach to environmental ethics in terms of naturalization of semantic information. According to him,

> There is value in the environment. This value consists in a certain sort of information, information that exists in the relation between affordances of the environment and their indices. This information exists independently of...sentient creatures. ...The information is there. It is in the world. What makes this information value, however, is the fact that it is valued by valuing creatures [because of evolutionary reasons], or that it would be valued by valuing creatures if there were any around. (p. 153)

Any form of life is deemed to enjoy some essential proprieties or moral interests that deserve and demand to be respected, at least minimally and relatively, that is, in a possibly overridable sense, when contrasted to other interests. So biocentric ethics argues that the nature and well being of the patient of any action constitute (at least partly) its moral standing and that the latter makes important claims on the interacting agent, claims that in principle ought to contribute to the guidance of the agent's ethical decisions and the constraint of the agent's moral behavior. The "receiver" of the action, the patient, is placed at the core of the ethical discourse, as a center of moral concern, while the "transmitter" of any moral action, the agent, is moved to its periphery.

Substitute now "life" with "existence" and it should become clear what IE amounts to. Information Ethics is an ecological ethics that replaces biocentrism with ontocentrism. It suggests that there is something even more elemental than life, namely, being—that is, the existence and flourishing of all entities and their global environment—and something more fundamental than suffering, namely, entropy. The latter is most emphatically not the physicists' concept of thermodynamic entropy. Entropy

here refers to any kind of destruction, corruption, pollution, and depletion of informational objects (mind, not of information as content), that is, any form of impoverishment of being. It is comparable to the metaphysical concept of nothingness. Information Ethics then provides a common vocabulary to understand the whole realm of being informationally. Information Ethics holds that being/information has an intrinsic worthiness. It substantiates this position by recognizing that any informational entity has a Spinozian right to persist in its own status, and a Constructionist right to flourish, i.e., to improve and enrich its existence and essence. As a consequence of such "rights," we shall see that IE evaluates the duty of any moral agent in terms of contribution to the growth of the infosphere and any process, action, or event that negatively affects the whole infosphere—not just an informational entity—as an increase in its level of entropy (or nothingness) and, hence, an instance of evil (Floridi and Sanders 1999, 2001; Floridi 2003).

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 center 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 (the collection of all the elementary proprieties that constitute it for what it is), which deserve to be respected (at least in a minimal and overridable sense) and, hence, place moral claims on the interacting agent and ought to contribute to the constraint and guidance of his ethical decisions and behavior. This ontological equality principle means that any form of reality (any instance of information/being), simply for the fact of being what it is, enjoys a minimal, initial, overridable, equal right to exist and develop in a way that is appropriate to its nature. The conscious recognition of the ontological equality principle presupposes a disinterested judgment of the moral situation from an objective perspective, i.e., a perspective which is as non-anthropocentric as possible. Moral behavior is less likely without this epistemic virtue. The application of the ontological equality principle is achieved whenever actions are impartial, universal, and "caring." At the roots of this approach lies the ontic trust binding agents and patients. A straightforward way of clarifying the concept of ontic trust is by drawing an analogy with the concept of "social contract."

Various forms of contractualism (in ethics) and contractarianism (in political philosophy) argue that moral obligation, the duty of political obedience, or the justice of social institutions gain their support from a so-called "social contract." This may be a hypothetical agreement between the parties constituting a society (e.g., the people and the sovereign, the members of a community, or the individual and the state). The parties accept to agree to the terms of the contract and thus obtain some rights in exchange for some freedoms that, allegedly, they would enjoy in a hypothetical state of nature. The rights and responsibilities of the parties subscribing to the agreement are the terms of the social contract, whereas the society, state, group, etc. is the entity created for the purpose of enforcing the agreement. Both rights and freedoms are not fixed and may vary, depending on the interpretation of the social contract.

Interpretations of the theory of the social contract tend to be highly (and often unknowingly) anthropocentric (the focus is only on human rational agents) and stress the coercive nature of the agreement. These two aspects are not characteristic of the concept of ontic trust, but the basic idea of a fundamental agreement between parties as a foundation of moral interactions is sensible. In the case of the ontic trust, it is transformed into a primeval, entirely hypothetical pact, logically predating the social contract, which all agents cannot but sign when they come into existence, and that is constantly renewed in successive generations.[22]

Generally speaking, a trust in the English legal system is an entity in which someone (the trustee) holds and manages the former assets of a person (the trustor, or donor) for the benefit of certain persons or entities (the beneficiaries). Strictly speaking, nobody owns the assets, since the trustor has donated them, the trustee has only legal ownership, and the beneficiary has only equitable ownership. Now, the logical form of this sort of agreement can be used to model the ontic trust in the following way:

- the assets or "corpus" is represented by the world, including all existing agents and patients;
- the donors are all past and current generations of agents;
- the trustees are all current individual agents; and
- the beneficiaries are all current and future individual agents and patients.

By coming into being, an agent is made possible thanks to the existence of other entities. It is therefore bound to all that already is both unwillingly and inescapably. It should be so also caringly. Unwillingly because no agent wills itself into existence, though every agent can, in theory, will itself out of it. Inescapably because the ontic bond may be broken by an agent only at the cost of ceasing to exist as an agent. Moral life does not begin with an act of freedom but it may end with one. Caringly because participation in reality by any entity, including an agent—that is, the fact that any entity is an expression of what exists—provides a right to existence and an invitation (not a duty) to respect and take care of other entities. The pact then involves no coercion, but a mutual relation of appreciation, gratitude, and care, which is fostered by the recognition of the dependence of all entities on each other. Existence begins with a gift, even if possibly an unwanted one. A fetus will be initially only a beneficiary of the world. Once she is born and has become a full moral agent, she will be, as an individual, both a beneficiary and a trustee of the world. She will be in charge of taking care of the world, and, insofar as she is a member of the generation of living agents, she will also be a donor of the world. Once dead, she will leave the world to other agents after her and thus become a member of the generation of donors. In short, the life of a human agent becomes a journey from being only a beneficiary to being only a donor, passing through the stage of being a responsible trustee of the world. We begin our career of moral agents as strangers to the world; we should end it as friends of the world.

The obligations and responsibilities imposed by the ontic trust will vary depending on circumstances but, fundamentally, the expectation is that actions will be taken or avoided in view of the welfare of the whole world.

The crucial importance of the radical change in ontological perspective cannot be overestimated. Bioethics and Environmental Ethics fail to achieve a level of complete impartiality because they are still biased against what is inanimate, lifeless, intangible, or abstract (even Land Ethics is biased against technology and artefacts, for example). From their perspective, only what is intuitively alive deserves to be

considered as a proper center of moral claims, no matter how minimal, so a whole universe escapes their attention. Now, this is precisely the fundamental limit overcome by IE, which further lowers the minimal condition that needs to be satisfied, in order to qualify as a center of moral concern, to the common factor shared by any entity, namely, its informational state. And since any form of being is in any case also a coherent body of information, to say that IE is infocentric is tantamount to interpreting it, correctly, as an ontocentric theory.

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 (Hepburn 1984). As Naess (1973) has maintained, "all things in the biosphere have an equal right to live and blossom." There seems to be no good reason not to adopt a higher and more inclusive, ontocentric perspective. Not only inanimate but also ideal, intangible, or intellectual objects can have a minimal degree of moral value, no matter how humble, and so be entitled to some respect. There is a famous passage, in one of Einstein's letters, that well summarizes this ontic perspective advocated by IE.

> Some five years prior to his death, Albert Einstein received a letter from a nineteen-year-old girl grieving over the loss of her younger sister. The young woman wished to know what the famous scientist might say to comfort her. On March 4, 1950, Einstein wrote to this young person: 'A human being is part of the whole, called by us 'universe', a part limited in time and space. He experiences himself, his thoughts and feelings, as something separated from the rest, a kind of optical delusion of his consciousness. This delusion is a kind of prison for us, restricting us to our personal desires and to affection for a few persons close to us. Our task must be to free ourselves from our prison by widening our circle of compassion to embrace all humanity and the whole of nature in its beauty. Nobody is capable of achieving this completely, but the striving for such achievement is in itself a part of the liberation and a foundation for inner security'. (Einstein 1954)

Deep Ecologists have already argued that inanimate things too can have some intrinsic value. And in a well-known article, White (1967) asked, "Do people have ethical obligations toward rocks?" and answered that "To almost all Americans, still saturated with ideas historically dominant in Christianity…the question makes no sense at all. If the time comes when to any considerable group of us such a question is no longer ridiculous, we may be on the verge of a change of value structures that will make possible measures to cope with the growing ecologic crisis. One hopes that there is enough time left." According to IE, this is the right ecological perspective and it makes perfect sense for any religious tradition (including the Judeo-Christian one) for which the whole universe is God's creation, is inhabited by the divine, and is a gift to humanity, of which the latter needs to take care. Information Ethics translates all this into informational terms. If something can be a moral patient, then its nature can be taken into consideration by a moral agent A, and contribute to shaping A's action, no matter how minimally. In more metaphysical terms, IE argues that all aspects and instances of being are worth some initial, perhaps minimal and overridable, form of moral respect.

Enlarging the conception of what can count as a center of moral respect has the advantage of enabling one to make sense of the innovative nature of ICT, as providing a new and powerful conceptual frame. It also enables one to deal more satisfactorily with the original character of some of its moral issues, by approaching them from a theoretically strong perspective. Through time, ethics has steadily moved from a narrow to a more inclusive concept of what can count as a center of moral worth, from the citizen to the biosphere (Nash 1989). The emergence of the infosphere, as a new environment in which human beings spend much of their lives, explains the need to enlarge further the conception of what can qualify as a moral patient. Information Ethics represents the most recent development in this ecumenical trend, a Platonist and ecological approach without a biocentric bias.

More than fifty years ago, Leopold defined Land Ethics as something that "changes the role of Homo sapiens from conqueror of the land-community to plain member and citizen of it. It implies respect for his fellow-members, and also respect for the community as such. The land ethic simply enlarges the boundaries of the community to include soils, waters, plants, and animals, or collectively: the land" (Leopold 1949, 403). Information Ethics translates environmental ethics into terms of infosphere and informational objects, for the land we inhabit is not just the earth.

6. Conclusion

As a consequence of the re-ontologization of our ordinary environment, we shall be living in an infosphere that will become increasingly synchronized (time), delocalized (space), and correlated (interactions). Previous revolutions (especially the agricultural and the industrial ones) created macroscopic transformation in our social structures and architectural environments, often without much foresight. The informational revolution is no less dramatic. We shall be in serious trouble if we do not take seriously the fact that we are constructing the new environment that will be inhabited by future generations (Floridi and Sanders 2005). We should be working on an ecology of the infosphere if we wish to avoid problems such as a tragedy of the digital commons (Greco and Floridi 2004). Unfortunately, I suspect it will take some time and a whole new kind of education and sensitivity to realize that the infosphere is a common space, which needs to be preserved to the advantage of all. One thing seems unquestionable, though: the digital divide will become a chasm, generating new forms of discrimination between those who can be denizens of the infosphere and those who cannot, between insiders and outsiders, between information rich and information poor. It will redesign the map of worldwide society, generating or widening generational, geographic, socio-economic, and cultural divides. But the gap will not be reducible to the distance between industrialized and developing countries, since it will cut across societies (Floridi 2002). We are preparing the ground for tomorrow's digital favelas.[23]

Endnotes

1. Source: Lyman and Varian (2003). An exabyte is approximately 10^{18} bytes, or a billion times a billion bytes.
2. Source: Computer Industry Almanac, Inc.
3. Source: U.S. Census Bureau's Statistical Abstract of the United States.
4. It is an aging population: the average game player is thirty-three years old and has been playing games for twelve years, while the average age of the most frequent game buyer is forty years old. The average adult woman plays games 7.4 hours per week. The average adult man plays 7.6 hours per week. Source: Entertainment Software Association, http://www.theesa.com/facts/top_10_facts.php
5. See, for example, Floridi (1999b) or http://en.wikipedia.org/wiki/Infosphere
6. These sections are based on Floridi (2006) and Floridi (2007b).

7. This section is based on Floridi (1999a), Floridi (2007a), and Floridi (forthcoming).
8. http://www.telegraph.co.uk/news/main.jhtml?xml=/news/2006/03/02/wkat02.xml&sSheet=/news/2006/03/02/ixworld.html and http://en.wikipedia.org/wiki/Tessa_Jowell_financial_allegations
9. I have analyzed this IT-heodicean problem and the tragedy of the good will in Floridi and Sanders (2001) and in Floridi (2006).
10. The term is introduced by Westin (1968) to describe a digital profile generated from data concerning a user's habits online.
11. The outer world, or reality, as it affects the agent inhabiting it.
12. For a similar position in computer ethics see Maner (1996). On the so-called "uniqueness debate" see Floridi and Sanders (2002a) and Tavani (2002).
13. The interested reader may find a detailed analysis of the model in Floridi (forthcoming).
14. The classic reference here is to Wiener (1954). Bynum (2001) has convincingly argued that Wiener may be considered one of the founding fathers of Information Ethics.
15. See Floridi and Sanders (2002b) for a defense of this position.
16. An early review is provided by Smith (1996).
17. For an analysis of the so-called IT-heodicean problem and of the tragedy of the good will, see Floridi (2006).
18. One may recognize in this approach to Information Ethics a position broadly defended by van den Hoven (1995) and more recently by Mathiesen (2004), who criticizes Floridi and Sanders (1999) and is in turn criticized by Mather (2005). Whereas van den Hoven purports to present his approach to IE as an enriching perspective contributing to the debate, Mathiesen means to present her view, restricted to the informational needs and states of the moral agent, as the only correct interpretation of IE. Her position is thus undermined by the problems affecting any microethical interpretation of IE, as Mather well argues.
19. For further details see Floridi (2005a).
20. For a detailed analysis and defense of an object-oriented modelling of informational entities see Floridi (1999a), Floridi and Sanders (1999), and Floridi (2003).
21. "Perspective" here really means level of abstraction; however, for the sake of simplicity the analysis of levels of abstractions has been omitted from this chapter. The interested reader may wish to consult Floridi (forthcoming).
22. There are important and profound ways of understanding this Ur-pact religiously, especially but not only in the Judeo-Christian tradition, where the parties involved are God and Israel or humanity, and their old or new covenant makes it easier to include environmental concerns and values otherwise overlooked from the strongly anthropocentric perspective prima facie endorsed by contemporary contractualism. However, it is not my intention to endorse or even draw on such sources. I am mentioning the point here in order to shed some light both on the origins of contractualism and on a possible way of understanding the onto-centric approach advocated by IE.
23. This paper is based on Floridi (1999a), Floridi and Sanders (2001), Floridi et al. (2003), Floridi and Sanders (2004b), Floridi (2005a), Floridi and Sanders (2005), Floridi (2006), Floridi (2007b), Floridi (2007a), and Floridi (forthcoming). I am in debt to all colleagues and friends who shared their comments on those papers. Their full list can be found in those publications. Here I wish to acknowledge that several improvements are due to their feedback. I am also very grateful to the editor, Peter Boltuć, for his kind invitation to contribute to this issue of the APA Newsletter on Philosophy and Computers.

References

Bynum, T. 2001. "Computer Ethics: Basic Concepts and Historical Overview." In The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta.

Einstein, A. 1954. Ideas and Opinions. New York: Crown Publishers.

Floridi, L. "Information Ethics: On the Philosophical Foundations of Computer Ethics." Ethics and Information Technology 1 (1999a): 37-56.

——. 1999b. Philosophy and Computing: An Introduction. London, New York: Routledge.

——. "Information Ethics: An Environmental Approach to the Digital Divide." Philosophy in the Contemporary World 9 (2002): 39-45.

——. "On the Intrinsic Value of Information Objects and the Infosphere." Ethics and Information Technology 4 (2003): 287-304.

——. 2005a. "A Model of Data and Semantic Information." In Knowledge in the New Technologies, edited by Gerassimos Kouzelis, Maria Pournari, Michael Stöppler, and Vasilis Tselfes. 21-42. Berlin: Peter Lang.

——. "Presence: From Epistemic Failure to Successful Observability." Presence: Teleoperators and Virtual Environments 14 (2005b): 656-67.

——. "Information Technologies and the Tragedy of the Good Will." Ethics and Information Technology 8 (2006): 253-62.

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

——. "A Look into the Future Impact of ICT on Our Lives." The Information Society 23 (2007b): 59-64.

——. "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, forthcoming.

Floridi, L., Greco, G.M., Paronitti, G., and Turilli, M. 2003. "Di Che Malattia Soffrono I Computer?" ReS http://www.enel.it/magazine/res/

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

Floridi, L., and Sanders, J.W. "Entropy as Evil in Information Ethics." Etica & Politica, special issue on Computer Ethics 1(1999).

——. "Artificial Evil and the Foundation of Computer Ethics." Ethics and Information Technology 3 (2001), 55-66.

——. "Computer Ethics: Mapping the Foundationalist Debate." Ethics and Information Technology 4 (2002b): 1-9.

——. 2004a. "The Method of Abstraction." In Yearbook of the Artificial. Nature, Culture and Technology. Models in Contemporary Sciences, edited by M. Negrotti. 177-220. Bern: Peter Lang.

——. "On the Morality of Artificial Agents." Minds and Machines 14 (2004b): 349-79.

——. 2005. "Internet Ethics: The Constructionist Values of Homo Poieticus." In The Impact of the Internet on Our Moral Lives, edited by Robert Cavalier. New York: SUNY.

Greco, G.M., and Floridi, L. "The Tragedy of the Digital Commons." Ethics and Information Technology 6 (2004): 73-82.

Hepburn, R. 1984. Wonder and Other Essays. Edinburgh: Edinburgh University Press.

Leopold, A. 1949. The Sand County Almanac. New York: Oxford University Press.

Lyman, P., and Varian, H.R. 2003. "How Much Information 2003."

Maner, W. "Unique Ethical Problems in Information Technology." Science and Engineering Ethics 2 (1996): 137-54.

Mather, K. "Object Oriented Goodness: A Response to Mathiesen's 'What Is Information Ethics?'" Computers and Society 34 (2005), http://www.computersandsociety.org/sigcas_ofthefuture2/sigcas/subpage/sub_page.cfm?article=919&page_number_nb=911

Mathiesen, K. "What Is Information Ethics?" Computers and Society 32 (2004), http://www.computersandsociety.org/sigcas_ofthefuture2/sigcas/subpage/sub_page.cfm?article=909&page_number_nb=901

Naess, A. "The Shallow and the Deep, Long-Range Ecology Movement," Inquiry 16 (1973): 95-100.

Nash, R.F. 1989. The Rights of Nature. Madison, Wisconsin: The University of Wisconsin Press.

Rawls, J. 1999. A Theory of Justice, Rev. Ed. Oxford: Oxford University Press.

Smith, M.M. 1996. "Information Ethics: An Hermeneutical Analysis of an Emerging Area in Applied Ethics," Ph.D. thesis, The University of North Carolina at Chapel Hill, Chapel Hill, NC.

Tavani, H.T. "The Uniqueness Debate in Computer Ethics: What Exactly Is at Issue, and Why Does It Matter?" Ethics and Information Technology 4 (2002): 37-54.

van den Hoven, J. 1995. "Equal Access and Social Justice: Information as a Primary Good," ETHICOMP95: An international conference on the ethical issues of using information technology. Leicester, UK: De Montfort University.

Westin, A.F. 1968. Privacy and Freedom 1st. New York: Atheneum.

White, L.J. "The Historical Roots of Our Ecological Crisis." Science 155 (1967): 1203-07.

Wiener, N. 1954. The Human Use of Human Beings: Cybernetics and Society, Rev. Ed. Boston: Houghton Mifflin.

Towards Artificial Consciousness

Riccardo Manzotti
IULM University

Abstract

In recent years, several researchers ventured the hypothesis of designing and implementing a model for artificial consciousness—there is hope of being able to design a model for artificial consciousness and of using such models for understanding human consciousness. The traditional field of Artificial Intelligence is now flanked by the seminal field of artificial or machine consciousness. In this paper, I analyze the current state of the art of models of consciousness and then present a model of consciousness based on a process-oriented view of consciousness. Eventually, I will sketch the relation between this model and the capability of developing new goals in an agent.

1. Consciousness and artificial consciousness

During the last ten years, interest in the scientific understanding of the nature of consciousness has been rekindled (Hameroff, Kaszniak, et al. 1996; Jennings 2000; Miller 2005). To date, a satisfactory and accepted framework has not been achieved either because experimental data is scarce or because a misleading theoretical standpoint is assumed.

The effort for a scientific understanding of consciousness has been flanked by a related approach named artificial consciousness (sometimes machine or synthetic consciousness) aiming at reproducing the relevant features of consciousness using non-biological components (Holland 2003; Adami 2006; Chella and Manzotti 2007). This new field has strong relationships with artificial intelligence and cognitive robotics.

Most mammals seem to show some kind of consciousness. It is highly probable that a conscious agent has some evolutionary advantage. Although it is still difficult to outline a precise functional role of consciousness, many believe that consciousness endorses a more robust autonomy, a higher resilience, a more general problem-solving capability, reflexivity, and self-awareness (Adami 2006; Bongard, Zykov, et al. 2006).

At the same time, trying to implement a conscious machine is a feasible approach to the scientific understanding of consciousness itself. Edelman and Tononi wrote that

> to understand the mental we may have to invent further ways of looking at brains. We may even have to synthesize artifacts resembling brains connected to bodily functions in order fully to understand those processes. Although the day when we shall be able to create such conscious artifacts is far off we may have to make them before we deeply understand the processes of thought itself. (Edelman and Tononi 2000)

According to Owen Holland (2003), it is possible to distinguish between Weak Artificial Consciousness and Strong Artificial Consciousness. Holland defines them as follows:

1) Weak Artificial Consciousness: design and construction of machines that simulate consciousness or cognitive processes usually correlated with consciousness.

2) Strong Artificial Consciousness: design and construction of conscious machines.

Most of the people currently working in the field of Artificial Consciousness would embrace the former definition. At any rate, the boundaries between the two are not easy to define. For instance, if a machine could exhibit all behaviors normally associated with a conscious being, would it be a conscious machine? Are there behaviors that are uniquely correlated with consciousness?

On the other hand, other authors claim that Artificial Consciousness could provide a better foundation for complex control whenever autonomy has to be achieved. In this respect, Artificial Consciousness could be, at least in principle, applied to all kinds of complex systems ranging from a petrochemical plant to a complex network of computers. The complexity of current artificial systems is such that outperforms traditional control techniques. Artificial consciousness could provide new ways to control. According to Ricardo Sanz, there are three motivations to pursue Artificial Consciousness (Sanz 2005; Bongard, Zykov, et al. 2006):

1) implementing and designing machines resembling human beings (cognitive robotics);
2) understanding the nature of consciousness (cognitive science);
3) implementing and designing more efficient control systems.

The most dreaded aspect of consciousness, which justifies this careful subdivision of the field, is the so-called "Hard-Problem" of consciousness. The label "Hard-Problem" of consciousness was coined by David Chalmers (1996), when he distinguished between "easy problems" of understanding consciousness (such as explaining the ability to discriminate, integrate information, report mental states, focus attention, etc.) and contrasted them with the "hard problem" (Why does awareness of sensory information exist at all? And why is there a subjective component to experience?). It is easy to see that the separation between Weak and Strong Artificial Consciousness mirrors the separation between the easy problems and the hard problem of consciousness.

One final dichotomy that is worth mentioning is the one between Access Consciousness and Phenomenal Consciousness (often abbreviated in A-Consciousness and P-Consciousness). According to a pivotal paper by Block (1995, 2002), there is confusion about the word consciousness regarding its twofold meaning—"the concept of consciousness is a hybrid or better, a mongrel concept." This confusion is dispelled distinguishing between access-consciousness and phenomenal consciousness.

A-consciousness is mostly functional whereas reportability and control are prominently important. According to Block, A-consciousness is definable as "poised for control of speech, reasoning and action" (Block 1995). Other authors, like David Chalmers, suggest defining A-consciousness "as directly available of global control" (Chalmers 1997).

On the contrary, P-consciousness is conscious experience; what makes a state phenomenally conscious is that there is something it is like to be in that state.

2. Internalism vs. externalism

Currently, the majority of scientific and philosophical literature on consciousness is biased by a seldom challenged assumption—the separation between the subject and the object. Although it is obvious that the body of the subject is separate from the body of the object, it is by no means so obvious that the mind is confined by the same boundaries of the brain. "Where does the mind stop and the rest of the world begin? …Someone accepts the demarcations of skin and skull, and say that what is outside the body is outside the mind" (Clark and Chalmers 1999). Indeed, there are many phenomena that extend beyond the boundaries of the body (behaviors, actions, perceptions, ecological processes). The mind could be one of them.

With regard to the nature of the mind, two very broad standpoints must be considered: internalism and externalism. The former states that our consciousness is identical (or correlated) to the processes, events, or states of affairs going on inside the boundary of our body (or brain). The latter affirms that our consciousness might depend partially or totally on the events, processes, or states of affairs outside our head or even outside our body.

Most current approaches to the problem of consciousness lean towards the internalist viewpoint (Crick 1994; Edelman and Tononi 2000; Metzinger 2000; Rees, Kreiman, et al. 2002; Crick and Koch 2003; Koch 2004). However, this approach raises several conundrums. If the mind is entirely located or dependent on events or states of affairs located inside the cranium, how can they represent events taking place in the external world? Consciousness appears to have properties that differ from anything taking place inside the cranium (Place 1956). Spurred on by common sense, literature has revealed a very strong impulse to "etherealize" or "cranialize" consciousness (Honderich 2000). The internalist perspective has consistently led to dualism and still promotes a physicalist version of dualism by endowing the brain (or a brain subset, the Neural Correlate of Consciousness) with the same role as the dualistic subject. Koch's recent book (2004, p. 87), endorses an unbiased internalist view with respect to consciousness and the brain: "The entire brain is sufficient for consciousness—it determines conscious sensations day in and day out. Identifying all of the brain with the NCC [Neural Correlate of Consciousness], however is not useful because likely a subset of brain matter will do."

On the other hand, many authors, like myself, have questioned the separation between subject and object—between representation and represented. They are looking for a different framework in which subject and object are two different perspectives on the same physical phenomenon. Their views could be labeled as some kind of externalism (Hurley 2001, 2006).

According to Mark Rowlands, there are two variants of externalism: content externalism and vehicle externalism. The former corresponds to the "idea that the semantic content of mental states that have it is often dependent on factors…that are external to the subject of that content" (Rowlands 2003, p. 5). The latter is more radical and affirms that "the structures and mechanisms that allow a creature to possess or undergo various mental states and processes are often structure and mechanisms that extend beyond the skin of that creature" (Rowlands 2003, p. 6).

In the following paragraphs, I will present a version of vehicle externalism (Manzotti and Tagliasco 2001; Manzotti 2003, 2005, 2006b, 2006a).

3. The Enlarged Mind: An Externalist Framework

In order to support vehicle externalism as a framework for artificial consciousness, I outline a framework in which the separation between the conscious perception of the world and the perceived physical world is not reconsidered.

The rationale is the following: the agent is conscious of those parts of the environment that produce effects due to the agent's body and neural structure. Objects, patterns, wholes are singled out by the agent's body. Consciousness is the way in which the environment is entangled in the behavioral history of the agent.

The rainbow is perhaps the best example in which there is no separation between the observed object/event and the observer. The rainbow is not a physical material object but, rather, a process that needs the interaction with the agent's body. If no observer were there, would the rainbow take place? No, it would not, because the light rays would continue their travel in space without interacting and, eventually, they would spread in the surrounding environment. On the contrary, if an observer were there, the converging rays of light would have hit his/her photoreceptors, and a fast but complex chain of physical processes would have continued from the retina to the cortical areas up to a point where the process corresponding to the rainbow would reach its end.

The rainbow is not a thing: it is a process, in which there is an entanglement between a physical condition and the agent's body. The light rays do not constitute a distinctive unity (the rainbow) unless and until they are embedded in a process. The occurrence of the rainbow depends not only on the presence of the physical conditions given above and the observer, but on a causal continuity between the two. This approach suggests a kind of direct realism based on the sameness between the physical process embedded in the perceived object and phenomenal experience itself.

I elsewhere proposed to call this process—which is constitutive of what there is and what we perceive—an onphene, derived from the Greek words ontos (what there is) and phenomenon (what appears) (Manzotti and Tagliasco 2001; Manzotti 2003, 2006a, 2006b). It refers to a process in which the traditional distinction between cause and effect (perceiver and perceived, representation and represented, subject and object) is missing.

The traditional problems of consciousness can thus be reconsidered once an externalist and process-based standpoint is adopted. The world in which each subject is living is no longer a private bubble of phenomenal experiences concocted by the brain. Each subject lives in and experiences the real world—the two being different descriptions of the same process. As conscious agents, we are part of a physical flow of processes possible due to our physical structure. I venture to suggest that these processes have the same properties of our own experiences as well as the same properties of the external world. Thereby, I suggest that these processes are identical with conscious experience.

According to the view presented here, the mind is identical with everything the subject is conscious of—everything being a process. Furthermore, the existence of what the mind is conscious of depends on the occurrence of those processes that are identical with the mind itself.

I consider the physical process that begins in the external world and ends in the brain as a unity. Such process occurs thanks to the brain, to the body, and to the surrounding environment. The rainbow is an excellent example of a process in which the act of observation, the observer, and the observed entity cannot be split; all occur jointly. They are different ways

to look at the same process. But the example of the rainbow, though a very compelling one, is not unique in leading to this conclusion. I propose that most perceived objects (if not all) have a structure analogous to that of the rainbow. The relevance of this argument lies in the fact that the brain is not self-sufficient with respect to mental events. The brain could thus be envisaged as the end part of a larger network of physical processes.

This is a view that could be considered a kind of radical externalism. The mind is literally and physically identical with a collection of processes spanning in time and space beyond the boundaries of the brain and the body. It is also a realist standpoint since it assumes that the experience of the world regards the world itself and not a mental representation of it.

4. Teleologically open systems and externalism

Is it possible to design and implement an architecture exploiting an externalist-oriented view? Or, more modestly, is it possible to derive some practical constraints from an externalist view? The common ground could be a teleological entanglement between the agent and the environment. The continuity outlined by the externalist view could be achieved by a system teleologically open.

I should begin with some words of caution concerning the scope and limits of this final section. The basic idea under development is that a conscious agent is a physical system that entangles the environment in its teleological history. As a result, a conscious agent is capable of deriving from its environment not just new categories and representations but also new goals. It is thus instructive to compare artificial systems deprived of consciousness with supposed conscious biological beings such as mammals. The hypothesis I make is that the latter will have a much higher degree of teleological openness—that is, the capability of acquiring new goals. Acquiring new goals should be the glue that keeps together newer and newer processes originating in the environment.

Current implementations of artificial systems focus on implementation of intelligent algorithms to achieve a fixed goal (or a fixed set of goals). Conscious subjects are capable of developing unpredictable and unexpected new goals.

Artificial systems are frequently designed with a fixed set of goals. Designers focus their efforts to find "how" those goals can be achieved. Learning is usually defined as a modification in agents' behavior: a modification driven by a goal. Various learning paradigms focus mostly on this modification of behavior. Supervised and unsupervised learning as well as reinforcement learning are valid examples (Sutton and Barto 1998, p. 3): they are based on fixed goals. For instance, in a reinforcement learning based agent "the reward function [corresponding to the goal] must necessarily be unalterable by the agent" (Sutton and Barto 1998, p. 8). On the contrary, many biological systems are capable of developing partially or totally unpredictable goals. There is evidence that such capability is greater in humans and mammals.

First, to develop new goals is important since the environment cannot be completely predicted at design time. Therefore, a truly adaptive system must be able to add new goals, not only to modify its behavior in order to perform optimally on the basis of some fixed criteria but also to change the criteria.

The behavior of behavior-based artificial agents depends on experience and goals defined elsewhere at design time (Arkin 1999). Motivation-based agents begin to show the capability of developing new goals (Manzotti and Tagliasco 2005). In complex biological systems, behavior depends on experience and goals; yet, goals are not fixed. Goals are the result of the interaction between the subject and its environment. In many complex biological systems, it is possible to distinguish between phylogenetic aspects and ontogenetic ones, nature versus nurture (Gould 1977; Elman, Bates, et al. 2001; Ridley 2004).

What is a goal? An agent's goal is an event whose occurrence is more probable thanks to the agent's structure (cognitive and bodily). Goals are embedded in causal structures that link the past with the future, the environment with the agent.

It is possible to classify artificial agents accordingly to their degree of teleological plasticity: fixed control architectures, learning architectures and goal-generating architectures. In the first case, the system has no capability of modifying how it does what it does. In the second case, the system is capable of modifying its behavior to fulfill some a priori target. The system is capable of modifying how it behaves. In the third case, the system is capable of modifying not only how it does what it does, but also what it does.

Systems with a fixed control architecture have a fixed causal structure. There is no ontogenesis whatsoever. Notwithstanding the behavioral complexity of the system, everything happens because it has been previously coded within the system structure. A mechanical device and a complex software agent are not different in this respect: both are pre-programmed in what they must achieve and how they must achieve it. Nothing in their structure is caused by their experiences. Suitable examples of this category are Tolam's artificial sow bug, Braitenberg's thinking vehicles (Braitenberg 1984), Brooks' artificial insects, and recent entertainment robots like Sony AIBO and Honda's humanoid ASIMO (2002).

A different level of dependency with the environment is provided by architectures that can learn how to perform a task. Behavior-based robots can be classified in this category. Systems based on artificial neural networks are well-known examples of this kind of architecture. These systems determine how to get a given result once they have been provided with a specific goal. The goal can be given either as a series of examples of correct behavior (supervised learning) or as a simple evaluation of the global performance of the system (reinforcement learning) (Sutton and Barto 1998). In both cases some kind of learning is applied. These systems lack the capability of creating new goals. By controlling its motors, a behavior-based robot can learn how to navigate avoiding static and dynamic obstacles. However, the goal behind this task is defined by the a priori design of the system. There are several examples of this kind of learning agent: Babybot at LIRA-Lab (Metta, Manzotti, et al. 2000) and Cog at MIT (Brooks, Breazeal, et al. 1999).

An agent, which learns both how to perform a given task and what task, corresponds to a teleologically open architecture. This is the case for most, if not all, mammals; it is true for primates and for human beings. They are systems capable of developing new goals that do not belong to their genetic background. For their development, these systems depend more on the environment than the previous two categories. A system belonging to the first category does not depend on the environment for what it does or for how it does what it does. A system belonging to the second category does depend on the environment for how it does what it does, but not for what it does. A system belonging to the third and last category depends on the environment both for what and how it does what it does.

I suggest that the kind of environmental entanglement achieved by a teleologically open architecture is the same exploited by human beings when they perceive consciously. It is a framework that could be used to deal with phenomenal consciousness in the field of artificial consciousness. If it could be proven to be correct, it would not require any kind of biological neural activity like those implicitly assumed by most

of the NCC-biased literature. Consciousness would be identical with the right kind of causal entanglement between an agent and its environment.

References

Adami, C. "What Do Robots Dream Of?" Science 314 (2006): 1093-94.

Arkin, R.C. 1999. Behavior-Based Robotics. Cambridge, MA: MIT Press.

Block, N. "On a Confusion about a Function of Consciousness." Behavioral and Brain Sciences 18 (1995): 227-87.

Block, N. 2002. "Some Concepts of Consciousness." In Philosophy of Mind: Classical and Contemporary Readings, edited by D. Chalmers. Oxford: Oxford University Press.

Bongard, J., V. Zykov, et al. "Resilient Machines through Continuous Self-Modeling." Science 314 (2006): 1118-21.

Braitenberg, V. 1984. Vehicles: Experiments in Synthetic Psychology. Cambridge, MA: MIT Press.

Brooks, R.A., C. Breazeal, et al. 1999. "The Cog Project: Building a Humanoid Robot." In Computation for Metaphors, Analogy, and Agents, edited by Nehaniv. Berlin, Springer-Verlag, 1562.

Chalmers, D.J. 1996. The Conscious Mind: In Search of a Fundamental Theory. New York: Oxford University Press.

Chalmers, D.J. 1997. "Availability: The Cognitive Basis of Experience." In The Nature of Consciousness, edited by N. Block, O. Flanagan, and G. Guzeldere. 421-23. Cambridge, MA: MIT Press.

Chella, A. and R. Manzotti. 2007. Artificial Consciousness. Exeter (UK): Imprint Academic.

Clark, A. and D. Chalmers. "The Extended Mind." Analysis 58 (1999): 10-23.

Crick, F. 1994. The Astonishing Hypothesis: the Scientific Search for the Soul. New York: Touchstone.

Crick, F. and C. Koch. "A framework for consciousness." Nature Neuroscience 6 (2003): 119-26.

Edelman, G.M. and G. Tononi. 2000. A Universe of Consciousness. How Matter Becomes Imagination. London: Allen Lane.

Elman, J.L., E.A. Bates, et al. 2001. Rethinking Innateness: A Connectionist Perspective on Development. Cambridge, MA: MIT Press.

Gould, S.J. 1977. Ontogeny and Phylogeny. Cambridge, MA: Harvard University Press.

Hameroff, S.R., A.W. Kaszniak, et al. 1996. Toward a Science of Consciousness: The First Tucson Discussions and Debates. Cambridge, MA: MIT Press.

Holland, O. 2003. Machine Consciousness. New York: Imprint Academic.

Honderich, T. "Consciousness as Existence Again." Theoria 95 (2000): 94-109.

Hurley, S.L. "Perception and Action: Alternative Views." Synthese 129 (2001): 3-40.

Hurley, S.L. 2006. "Varieties of externalism." In The Extended Mind, edited by R. Menary. Aldershot: Ashgate Publishing.

Jennings, C. "In Search of Consciousness." Nature Neuroscience 3 (2000): 1.

Koch, C. 2004. The Quest for Consciousness: A Neurobiological Approach. Englewood, CO: Roberts & Company Publishers.

Manzotti, R. 2003. "A Process Based Architecture for an Artificial Conscious Being." In Process Theories, edited by J. Seibt. 285-312. Dordrecht: Kluwer Academic Press.

Manzotti, R. 2005. "The What Problem: Can a Theory of Consciousness be Useful?" In Yearbook of the Artificial, edited by P. Lang. Berna.

Manzotti, R. "An Alternative Process View of Conscious Perception." Journal of Consciousness Studies 13 (2006a): 45-79.

Manzotti, R. "Consciousness and Existence as a Process." Mind and Matter 4 (2006b): 7-43.

Manzotti, R. and V. Tagliasco. 2001. Coscienza e Realtà. Una teoria della coscienza per costruttori e studiosi di menti e cervelli. Bologna: Il Mulino.

Manzotti, R. and V. Tagliasco. "From 'behaviour-based' robots to 'motivations-based' robots." Robotics and Autonomous Systems 51 (2005): 175-90.

Metta, G., R. Manzotti, et al. 2000. "Development: is it the right way towards humanoid robotics?" In ISA-6, Venezia: IOS Press.

Metzinger, T. 2000. Neural Correlates of Consciousness: Empirical and Conceptual Questions. Cambridge, MA: MIT Press.

Miller, G. "What are the Biological Basis of Consciousness?" Science 309 (July 2005): 79.

Place, U.T. "Is Consciousness a Brain Process?" British Journal of Psychology 47 (1956): 44-50.

Rees, G., G. Kreiman, et al. "Neural Correlates of Consciousness in Humans." Nature Reviews 3 (2002): 261-70.

Ridley, M. 2004. Nature via Nurture. Great Britain: Harper.

Rowlands, M. 2003. Externalism. Putting Mind and World Back Together Again. Chesham: Acumen Publishing Limited.

Sanz, R. 2005. "Design and Implementation of an Artificial Conscious Machine." In IWAC2005, Agrigento.

Sutton, R.S. and A.G. Barto. 1998. Reinforcement Learning. Cambridge, MA: MIT Press.

COMMENTARIES ON HARMAN

Formulating the Explanatory Gap

Yujin Nagasawa
University of Birmingham

Gilbert Harman (2007) purports to illuminate the intractability of the so-called "explanatory gap" between the phenomenal aspect of consciousness and an objective physical explanation of that aspect by constructing a parallel situation involving translation from one language to another. While I agree with several points that Harman makes regarding the nature of phenomenal consciousness, I have a reservation about his formulation of the explanatory gap. In what follows, I explain my reservation.

Harman's formulation is based on Thomas Nagel's well-known example of a bat, which Harman describes as follows:

> Nagel observes that there may be no such translation from certain aspects of the other creature's experiences into possible aspects of one's own experiences. As a result, it may be impossible for a human being to understand what it is like to be a bat.

Harman then explains the structure of a possible translation that would fill the explanatory gap:

> Suppose we have a completely objective account of translation from the possible experiences of one creature to those of another, an account in terms of objective functional relations, for example. That can 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 analogous requirements. First, one must be able to identify one objectively described conceptual system as one's own. Second, one must have in that system something with the same or similar functional properties as the given experience. To understand what it is like for the other creature to have that experience is to understand which possible experience of one's own is its translation.

Harman's description of the explanatory gap in terms of translation from bat experience to human experience seems to face the same problem that Nagel's description faces.

Nagel contends that it is difficult to know how physicalism could be true given that we cannot know what it is like to be a bat, or, that is, that we cannot know the phenomenal aspects of a bat's sensory experiences. Nagel's bat example is often said to be so effective because, to any intelligent person, it seems so obvious that a bat's sonar is nothing like any sensory apparatus that we have.

But exactly why does a bat's having a unique sensory apparatus make it impossible to know what it is like to be one? There are two possible explanations here:

(1) We have to be bats, or at least bat-type creatures that use sonar, in order to know what it is like to be a bat. However, we are neither bats nor bat-type creatures.

(2) An objective, physical characterization of a bat does not tell us what it is like to have sonar, and hence what it is like to be a bat.

Consider (1). If (1) is true, it is difficult to see why physicalism is threatened by the fact that we non-bats cannot know what it is like to be a bat. While physicalism is the ontological thesis that, roughly speaking, everything in this world is physical in the relevant sense, (1) does not entail any significant ontological claim that could undermine physicalism or indeed any other alternatives. It implies only that no human theory, whether it is based on physicalism, dualism, or neutral monism, can tell us what it is like to be a bat, merely because human beings are neither bats nor bat-type creatures. Hence, if (1) is the basis of Nagel's bat example, it is irrelevant to the cogency of physicalism.[1]

Consider (2). If Nagel and Harman rely on this explanation, then, while (2) is relevant to the cogency of physicalism, ironically, the apparent vividness of the bat example and Harman's illustration about a translation turn out to be irrelevant. For the plausibility of (2) remains the same even if we replace the term "bat" with, for example, "human being." We know perfectly well what it is like to be a human being subjectively, but we have no idea how to characterize it fully objectively and physically. This in itself creates the explanatory gap between the phenomenal aspect of consciousness and an objective physical explanation of that aspect.

The explanatory gap is a very general problem about characterizing fully objectively and physically the phenomenal aspect of consciousness. Thus, it does not really matter whether the phenomenal aspect in question is related to our own type of experience or to those of other animals. It is therefore misleading to say that the explanatory gap is a result of our lacking "a completely objective account of translation from the possible experiences of one creature to those of another." It is a problem of there being no completely objective account of any experience, whether it is bat experience or human experience.

Suppose we discover somehow that, surprisingly, there is a one-to-one correspondence between a bat's phenomenal experiences and a human being's experiences, and that what it is like to be a bat is identical to what it is like to be a human being. Alternatively, suppose that we are the only conscious creatures in the whole universe. The explanatory gap nevertheless remains unfilled because, again, we still do not know how to characterize fully objectively and physically what it is like to be a human being.

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.[2]

Endnotes

1. See Nagasawa (2004 and forthcoming) for related points.
2. I would like to thank Kaitlyn Patia for helpful comments.

References

Harman, Gilbert. "Explaining an Explanatory Gap." American Philosophical Association Newsletter on Philosophy and Computers 6 (2007).

Nagel, Thomas. "What is it Like to Be a Bat?" Philosophical Review 83 (1974): 435-50.

Nagasawa, Yujin. "Thomas vs. Thomas: A New Approach to Nagel's Bat Argument." Inquiry 46 (2004): 377-94.

Nagasawa, Yujin (forthcoming). God and Phenomenal Consciousness. New York: Cambridge University Press.

To Understand the Understanding—das Verstehen zu Verstehen: A Discussion of Harman's "Explaining an Explanatory Gap"

Marion Ledwig
University of Nevada–Las Vegas

Harman's main claim is that:

> a purely objective account of conscious experience cannot always by itself give an understanding of what it is like to have that experience. There will at least sometimes be an explanatory gap. This explanatory gap has no obvious metaphysical implications. It reflects the distinction between two kinds of understanding: objective understanding and Das Verstehen. (Harman 2007, p. 3)

While Harman, the Stuart Professor of Philosophy at Princeton University, has not given an argument in this very short article as to why this explanatory gap has no metaphysical implications, and a creationist would probably jump at the opportunity to find another explanatory gap for God to fill, I will concentrate on the view for which he has argued. First of all, that this explanatory gap reflects the distinction between two kinds of understanding, namely, between objective understanding and das Verstehen, depends on what Harman means by objective understanding and das Verstehen.

Harman (1999, p. 264) makes his account of Verstehen explicit: "The theory of Das Verstehen says that certain aspects of psychological and social phenomena can only be understood by imaginatively putting oneself in the other person's position." Yet, in this regard 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? For 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 American English has an expression that just applies to one sex; that is, one just says with regard to a very good-looking woman that she is gorgeous, but not with regard to a very good-looking man. And one would think, in order to completely comprehend or understand the meaning of a term, that one would also like to know why it is the case that, in American English, there are words such as "gorgeous" that apply just to one sex. In this regard, not only a historical explanation might be helpful, but also a cultural one.

Although Harman (1999, p. 274) admits: "Simply learning what the rules are for the use of an expression (or for the use of the concept expressed) does not always bring an understanding of meaning with it," and although Harman (1999, p. 275) claims that "[o]ne way to understand intuitionistic connectives is to learn to use them in such a way that this use is second nature. … The usage has to be internalized," it is not obvious to me what second nature and internalization include. Do historical and cultural explanations for a given usage also fall under them? Also, in this regard, the example he provides is of no further help: with regard to the term "quantum number" in quantum physics, which Harman considers a term he has not yet mastered, Harman (1999, p. 268) claims: "The explanation works only if I learn how to use the phrase 'quantum number' as physicists use it." Of course, one could claim a physicist might not need an account of where the term "quantum number" historically comes from in order to comprehend it; yet I think in my "gorgeous" case, a historical or cultural explanation of the term would have helped me to understand its usage fully.

With regard to das Verstehen, it was also not clear to me if Harman would allow for partial Verstehen, or whether Verstehen always has to be complete. From the following passage, one gets the impression that Verstehen is either given or not in Harman (2007, p. 3): "This is on the assumption that one has an expression 'E' in one's own language that correctly translates the expression in the other language. If not, Das Verstehen will fail." Yet, my "gorgeous" case seems to suggest that partial Verstehen is possible.

Another question I have with regard to Harman's account is whether the explanatory gap is inevitable. I got the impression that he thinks it is. But, in my opinion, this doesn't have to be the case. That is, I think it might be possible to close the gap over time. In order to make that more probable, we shouldn't look at the example of bats and what it feels like to be a bat, but perhaps at an idea that is much closer for us: how it feels to be a member of the opposite sex. And that is something we might get much closer to now. For on the Internet, one can take on any kind of identity, and indeed it has happened that men have taken on the identities of women and vice versa. In that way, one might at least partially experience what it is to be of the opposite sex because if one has taken on a man's identity, one can listen to men's talk, take part in their activities, and will also be treated like a man. Perhaps in that way and after long exposure, one will finally also know what it feels like to be a man. In order to get a complete idea of what it feels like to be a man or a woman, one can even have a sex-change operation. Perhaps, one might want to object, it still is not 100 percent fool-proof that the person then really feels like a member of the opposite sex, but I have at least made it appear much more plausible that this is possible, and that is all I intended to show here.

Harman comes up with the following thought experiment in order to make clear that an objective description of certain actions cannot yield a subjective understanding of the respective actions, so that there is an explanatory gap:

> Suppose, for example, we discover the following regularity in the behavior of members of a particular social group. Every morning at the same time each member of the group performs a fixed sequence of actions: first standing on tip toe, then turning east while rapidly raising his or her arms, then turning north while looking down, and so on, all this for several minutes. We can certainly discover that there is this objective regularity and be able accurately to predict that these people will repeat it every morning, without having any subjective understanding of what they are doing—without knowing whether it is a moderate form of calisthenics, a religious ritual, a dance, or something else. Subjectively to understand what they are doing, we have to know what meaning their actions have for them. That is, not just to see the actions as instances of an objective regularity. (Harman 2007, p. 2)

Yet, in my opinion, we could narrow down by means of experiments which meaning is the appropriate one for these movements. For instance, if it were just an ordinary dance, it wouldn't matter in which particular direction they would be turning, so one could place them in a room where they are not able to discern where east and north are, and ask them to perform their movements. On the supposition of an ordinary dance, it wouldn't matter where east and north are. So if they continued their movements, this interpretation would seem plausible. To distinguish a dance from calisthenics, one would assume that music is essential for a dance, whereas this doesn't have to be the case with regard to calisthenics. So if they never played music with the movements, one wouldn't consider that to be a dance.

If the movements, however, were something of importance to them, one could try to hinder the people from performing their movements and see their reactions. Whether it makes them sad or angry, if they are hindered and what they are willing to do to keep up their movements could signify something about the meaning of these movements. Also, if it were a religious ritual, they would probably object if anybody just joined them performing the same kinds of movements, because in order to belong to a certain religion, one usually has to go through some kind of initiation rite, such as baptism. Of course, it is much easier to ask them what they are doing instead of trying to eliminate all possible explanations by means of experiments, for the number of possible explanations is actually enormous, if not infinite. Yet, it doesn't seem impossible to narrow down the meaning of these movements by looking at plausible explanations of them and trying to eliminate the respective explanations one by one through means of experiments or even by an experimentum crucis in order to obtain their meaning. So from a third person perspective, and not only from a first person perspective, one can subjectively understand what they are doing by just looking and experimenting with which kinds of actions really are part of the objective regularity. In the case of aliens or artificial life forms, though, it might be much more difficult to determine the meaning from a third person perspective, just for the simple reason that what has meaning for them and what kind of meaning things have might differ significantly from what is the case with regard to humans.

Of course, Harman could reply that by experimenting, we try to determine what meaning the actions have for them, so that in the end Harman is right: in order to subjectively understand what they are doing, we have to know what meaning their actions have for them. Yet, in order to determine the meaning, one can both employ a first- or a third-person perspective. Moreover, sometimes a God's eye perspective might be quite helpful because it places the actions in a bigger context. That is, it would make quite a difference in meaning to have opened a door to a Jew, a Gypsy, or to any person of the other persecuted groups in Nazi-Germany, as opposed to opening a door to a Jew, a Gypsy, etc. in contemporary twenty-first-century Germany.

Yet this thought experiment isn't Harman's only defense of the objective-subjective distinction. Here is another one:

> With respect to pain and other sensory experiences there is a contrast between an objective understanding and a subjective understanding of what it is like to have that experience, where such a subjective understanding involves seeing how the objective

> experience as described from the outside translates into an experience one understands from the inside. (Harman 2007, p. 2)

Yet, if color just referred to how light is reflected from different surfaces, I am sure one could enhance the texture of these surfaces to the blind, so that he or she might touch the different enhanced surfaces and might get a better understanding what the differences in color refer to; in this way, even the blind could get an experience he or she understands from the inside. I admit that it is not exactly the same kind of experience, but a similar one. Hence, I would like to find ways to close the gap. Of course, this seems difficult to accomplish with regard to pain, but actually, although a person without pain receptors might not experience bodily pain, he or she might experience psychological pain, such as humiliation or when a beloved friend or relative dies. So there is a psychological equivalent to bodily pain that the person without pain receptors can take as a model of comparison for bodily pain.

Harman emphasizes that

> To use an objective account of translation to understand an expression as used in another language, at least two further things are required. First, one must be able to identify a certain objectively described language as one's own language. …Second, one must have in one's own language some expression that is used in something like the same way as the expression in the other language. (Harman 2007, p. 3)

Yet, with regard to the first point, the identification might be difficult to make, for languages change and actually comprise an enormous amount of words, not to mention grammar rules. The German of the twenty-first century differs from the German of the nineteenth century. Moreover, the Swiss German of the twenty-first century differs from the so-called High German in Germany of the twenty-first century, which are points of which Harman (1999, p. 267) is quite clearly aware. So where is the point to be made when one is justified in claiming that one has identified one's own language? That seems to me quite arbitrary to a certain extent, so that vagueness enters the scene. With regard to the second point, it would be nice to have a more precise formulation. What does "something like the same way" mean? Does my "gorgeous" case fall under this condition or not? That is, we have a word in German that is the translation of the English word "gorgeous," but one can apply it to both sexes. So does this word fulfill Harman's second point? That is unclear to me.

Additionally, Harman (2007, p. 3) applies "these thoughts about language to the more general problem of understanding what it is like for another creature to have a certain experience. …First, one must be able to identify one objectively described conceptual system as one's own." As before, I think that this identification might be difficult depending on how large one's own conceptual system actually is; also, how one is presented with that system might make it easier or more difficult to identify it as one's own.

Finally, there is also something on which I agree with Harman—although I think that his works are very stimulating—namely, that Nemirov (1980), Lewis (1998), and Jackson (2004) expose a bizarre view when they claim that "understanding what it is like to have a given experience is not an instance of knowing that something is the case" (Harman 2007, p. 3).

References

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

Harman, G. "Explaining an Explanatory Gap." APA Newletters 06 (Spring 2007): 2-3.

Jackson, F. 2004. "Mind and Illusion." P. Ludlow, Y. Nagasawa, and D. Stoljar. In There's Something about Mary: Essays on Phenomenal Consciousness and Frank Jackson's Knowledge Argument. Cambridge, MA: MIT Press.

Lewis, D. 1988. Proceedings of the Russellian Society, edited by J. Copley-Coltheart. 13: 29-57.

Nemirov, L. "Review of T. Nagel, Moral Questions." Philosophical Review 89 (1980): 475-76.

Discussion Papers

Computing and Philosophy: In Search of a New Agenda

Gaetano Aurelio Lanzarone
Università degli Studi dell'Insubria, Varese, Italy

This paper proposes a step forward with respect to the current state of the debate in computing and philosophy, by taking into account (relatively) recent advances in computer science not much considered up to now. Two examples are briefly discussed: Computational Reflection and Second Life. A relationship between these apparently very distant topics is also sketched.

After Alan Turing, not many computer scientists have been involved with the philosophical implications of developments in their field.[1] On the other hand, a remarkable number of philosophers have been attracted by philosophic questions arising from the computing area. In speculating about them, however, more often than not they were not patient enough to cope with the computer's intricacies,[2] where the very nature of this artifact actually hides.[3] Moreover, the debate seems to be stuck in the theoretical results of the thirties (in logic and foundations of computability theory) and in the picture of the computer as an isolated machine, as it was in the sixties. Many advances and insights developed in computing in the last decades have not yet entered into the discussion.

As a first example, take a revolutionary concept that has been investigated since 1980[4] and whose profound potential implications do not seem to have been perceived by philosophers. This is the concept of "computational reflection," which consists of the possibility of formalizing domain descriptions at different levels (e.g., an object level and one or more meta-levels) and of dynamically shifting the reasoning up and down among them.[5]

A lot of authoritative papers have been written in the last decades to discuss Turing's mathematical objection, i.e., the significance of Gödel's theorems on the limits of a formal system with respect to the theoretical possibility of developing an artificial intelligence comparable with human intelligence.[6] Among those who have participated in this discussion, it is common knowledge that Gödel's theorems are only valid within the standard logical setting of a fixed set of axioms (the "system") and that it is always possible to go beyond by enhancing the system. But this is viewed as just moving the limit a little forward and incurring in the infinite regression problem. Since then, however, the research has evolved, opening new possibilities.

After Russell's discovery of a contradiction in Frege's Begriffsschrift system, classical logic languages have been separated into first-order, second-order … levels. Gödel's theorems sanctioned the limits of mixing levels so as to make self-reference possible. After recovering from this

shock, however, logicians began to recognize that, in order to overcome the impossibility for a logic system to state its own properties (i.e., completeness and consistency), it is not necessary to completely separate levels and meta-levels. It is sufficient for meta-concepts (such as truth and theoremhood) not to be fully represented within the object-level system: useful partial forms of self-reference, or introspection, may be allowed, given enough attention to the technical details so as to avoid the evils of paradox and inconsistency (Perlis 1985).[7]

Feferman made a step forward in going beyond the concept of truth in a hierarchy of levels. He introduced the concept of a reflection principle, defined as:

> a description of a procedure for adding to any set of axioms A certain new axioms whose validity follow from the validity of the axioms A and which formally express, within the language of A, evident consequences of the assumption that all theorems of A are valid. …In contrast to an arbitrary procedure for moving from A_k to A_{k+1}, a reflection principle provides that the axioms of A_{k+1} shall express a certain trust in the system of axioms A_k (under suitable conditions). (Feferman 1962)

Reflection principles were also introduced, as mentioned above, in the computational setting, particularly in some areas of Artificial Intelligence, such as knowledge representation, automatic reasoning, and functional and logic programming languages. Especially in the logic programming community, there was long and fruitful research work and debate about keeping the object- and meta-levels separated or leaving them to interplay; both approaches have been developed and compared.[8] In the Artificial Intelligence community the advantage of structuring knowledge at different but interacting levels of abstraction has been widely recognized.[9] Computational reflection allows the dynamic interweaving of knowledge and meta-knowledge, reasoning and meta-reasoning, and also updating knowledge (thus learning) during reasoning.

I believe that such a harvest of scientific results in the logic and computing areas puts some philosophical topics in an entirely new perspective. The capability of jumping in and out of the "system" (equivalently, of being observer or part of the observed system) seems to be one of the most relevant and flexible capabilities of the human mind. How computational reflection can be put to work in artificial systems to imitate this capability is still to be explored. While technically several prototype implementations exist (with their ways of not being trapped in the infinite regression problem), philosophically the topic awaits further investigation.

A second example of novel developments in the computing field is suggested by recent news. The Swedish government announced having opened an embassy in Second Life. This virtual world is, in my view, a breakthrough. We have become somewhat acquainted with virtual reality, but we were careful to maintain a border between virtual and "real" reality. Might it be the case that we will have to revise this position, considering what is happening in Second Life?

As defined in its website, Second Life (SL) is a 3-D online digital world imagined, created and owned by its residents (starting in 2003). Residents are users, represented by their digital image called an avatar. At the time of this writing, it is reported that SL has 6,574,270 residents; 1,086,106 of them logged-in during the last thirty days.

In SL, virtual dollars can be converted to real dollars. Residents work at professions, either the same or different ones from what they do in Real Life (RL), and they sell their products and services. Artists, designers, and architects have begun to perform and display their work in SL and, if they were successful, have continued in RL. Some have followed the opposite path, while others display their work in both worlds at the same time. Swedish designers have reproposed in RL the forms of objects created by SL residents. There are stores in SL that display and sell various avatar models; at New York's Columbia University, an exhibition of the best avatars was held recently. There are SL journalists, reporters, photographers, fashion magazines, tourist agencies, and much more.

New jobs have been created in SL and there are people who earn their living by working only in SL. A resident has reached the first million (real) dollars selling property (land and houses) located in SL. Multinational companies (e.g., IBM, Reuters) are colonizing SL by opening their headquarters there. Most of all, events are created in SL, like the famous press conferences held by Dell and Sun Microsystems. In October 2006, the "Stand Up Against Poverty" campaign was organized in SL to focus on the problems of (real) world poverty. More and more universities hold classes on the SL campus.

Second Life contains various micro-worlds, where role-playing games can take place: it's as if SL were a real world and one went to Disneyland. One can also interact with other virtual worlds outside of SL, within that giant virtual super-world called Internet. Blogs exist that comment on SL locations. An organization that has opened a headquarters in SL can display at a stand there its promotional material that points to its traditional website, or vice versa. A combination of SL and Skype allows in-world (as SL residents are called) people to interact at the same time also by phone.[10]

Second Life and Real Life thus combine and interact in various ways. In this respect, SL is not a usual website. We are acquainted with websites, of a company, a museum, etc., that refer to the corresponding real company, museum, etc.: these are first-order virtual reality (VR) environments (and a meta-level with respect to RL). Most of the places populated in SL, and the activities and events taking place there, have no such correspondence. At any moment, a user may just watch (on the screen of his computer) what happens in SL, or he may participate (through his avatar) in an activity, e.g., playing with a slot-machine. In the latter case, the user watches, on his computer, himself playing with the slot-machine. SL is therefore a second-order virtual reality environment (or a meta-metalevel, if the metalevel is the first-order VR).

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 interaction between RL and SL.

Up to this point, I have introduced Computational Reflection and Second Life just as two examples of advances in computing that have not entered into the philosophical discussion. They are very different from each other: on the one hand, Computational Reflection is an example of novel theoretical results (beyond those of the thirties), of highly speculative interest to researchers but (presently) scarcely known outside their community; on the other hand, Second Life is an example of what a profound impact can result from millions of interconnected computers (differently from the situation of the sixties), and what an influence it can have on a multitude of normal people.

With a certain degree of bravery, a parallel might even be drawn between these two new realities. An aspect that they have in common is that both are related to the topic of the inside (entering into) and the outside (exiting from) of a system.

The system's borders are dynamic, and the roles of observer/observed are interchangeable. In a sense, the reciprocal mirroring and interplay of levels that has been investigated in advanced logic and computing is now being "incarnated" in virtual worlds like SL.

The purpose of the present paper is to argue that it is time to take a step forward from what the epistemological discussion in computing and philosophy has already achieved. In my view, this will probably need a stronger interaction between philosophers and computer scientists, provided that the latter wish to attempt communicating their research results in a less technicality-driven way and the former wish to attempt giving a closer look at scientific findings. In any case, I would like to solicit a discussion based on the most advanced features, on the cutting edge of computer science research.

At this point in the time-line of the debate, we can take for granted that to ascribe intelligence to an (either natural or artificial) agent, this has to prove to be situated-in-a-context, reactive toward the environment, and capable of entertaining rich interactions with other agents. No intelligence can be born and can blossom in isolation, and computers have lived in isolation for about the first half of their life. But this no longer holds: nobody can conceive nowadays of a computer not connected to Internet, and this gives the computer an entirely new dimension. The Turing test is less at home in the Loebner Prize than it is vividly present in Second Life, especially when artificial agents will interact with human users.

In this paper, I have touched upon the topic of organizing descriptions of reality at different levels of languages and theories, and of providing the means to let the levels dynamically interact with each other in a coherent, but non-reducible, fashion. Further philosophical investigation[12] could clarify if this might lead to a third way between the approach of leaving different levels of abstraction completely separated (or only connected by static interfaces) and the "reductionist" approach of substituting higher-level propositions for lower-level propositions.

Self-reference and introspection appear to be a fundamental characteristic of complex systems, as pointed out for example by von Foerster (1981). It shows up in a variety of social phenomena; for instance, Luhmann (1996) has analyzed how today's mass media react primarily to themselves and only secondarily to the outside world. Self-reference is also a basic feature of natural languages, and thus of the human mind.

In earlier work, recursion has been viewed as the computational form of introspection, which in turn is the basis of self-awareness and consciousness, as expressed for instance by Nelson:

> Mechanism is the philosophy that the human mind is an information processing system. My own version of it says that mind is a system of recursive rules…complex enough to account for intentional attitudes such as belief and desire, and capable of sentience and self-awareness. (Nelson 1982)

Recursion is usually defined and considered at a single level of language (and of a theory expressed in that language). I mentioned above self-reference with reflection as recursion through levels of languages and theories. This seems to me a deeper insight into the capabilities of computer-based "intelligent" systems and of humans viewed as information processing systems, an insight that appears quite close to the intuition expressed by Hofstadter:

> My belief is that the explanation of emergent phenomena in our brains…for instance ideas, hopes, images, analogies, and finally consciousness and free will…are based on a kind of Strange Loop, an interaction between levels in which the top level reaches back towards the bottom level and influences it, while at the same time being itself determined by the bottom level. (Hofstadter 1979, p. 709)

Computing results are beginning to substantiate this intuition and deserve a closer philosophical consideration.

Endnotes

1. Exceptions are, e.g., Weizenbaum 1976, Winograd and Flores 1988.
2. Again, there are exceptions. For instance, Donald Gillies wrote a book (1996) after spending several months working together with Artificial Intelligence researchers.
3. As Dijkstra pointed out many years ago, the computer, with all its layers understood as a hierarchy of abstract machines, is the most complex machine ever created by mankind and is the artifact closest to the complexity of the human mind.
4. The seminal works were Weyhrauch 1980 and Smith 1984.
5. An overview of the procedural, logical, functional, and object-oriented approaches to computational reflection is, e.g., in Demers and Malenfant 1995. A fully worked out computational logic approach is in Barklund et al. 2000 and some epistemological remarks arising thereof are in Lanzarone 2003.
6. I skip the references to, e.g., Lucas', Searle's, et al. papers, replies, and counter-replies.
7. I avoid here more rigorous, but more cumbersome, definitions of first-order/higher-order languages, object/meta levels, static addition of axioms vs. dynamic application of reflection principles, self-reference/introspection, etc.
8. The former approach was followed in Hill et al. 1994. The latter approach was first introduced by Bowen and Kowalski 1982 and then pursued, among others, in Costantini and Lanzarone 1994. This dichotomy is, however, a simplification, as there is a wide range of intermediate or other solutions.
9. See, e.g., Carlucci Aiello and Levi 1988, Genesereth 1987, among many others.
10. My reference here is Gerosa 2007. This book, in Italian, is the first and only one, to my knowledge, that has made an analysis of how life is in SL, with interviews of successful residents. The books published in English are mainly technical manuals.
11. As a matter of fact, Neal Stephenson's novel Snow Crash, the precursor of all virtual worlds, introduced the term metaverse.
12. For instance, distinguishing between epistemological and ontological levels, or other forms of "levellism," along the lines of Floridi 2004.

Bibliography

Barklund J., Costantini S., Dell'Acqua P. and Lanzarone G.A. "Reflection Principles in Computational Logic." In Journal of Logic and Computation, volume 10, issue 6. 743-86. Oxford University Press, 2000.

Carlucci Aiello, L. and Levi, G. The Uses of Metaknowledge in AI Systems. In Meta-Level Architectures and Reflection, edited by P. Maes and D. Nardi. 243-54. North-Holland, Amsterdam, 1988.

Bowen, K.A. and Kowalski, R.A. "Amalgamating Language and Metalanguage." In Logic Programming, edited by K.L. Clark and S.A. Tarnlund. 153-72. Academic Press, 1982.

Costantini S. and Lanzarone G.A. "A Metalogic Programming Approach: Language, Semantics and Applications." International Journal of Experimental and Theoretical Artificial Intelligence 6 (1994): 239-87.

Demers F.N. and Malenfant J. "Reflection in Logic, Functional and Object-oriented Programming: a Short Comparative Study." In Procs. Fourteenth International Joint Conference on Artificial Intelligence, 1995.

Feferman S. "Transfinite Recursive Progressions of Axiomatic Theories." Journal of Symbolic Logic 27 (1962): 259-316.

Floridi L. and Sanders J.W., Levellism and the Method of Abstraction. IEG Research Report 22.11.2004, digital editing by G.M. Greco, Information Ethics Group, Oxford University – University of Bari, http://web.comlab.ox.ac.uk/oucl/research/areas/ieg

Genesereth, M.R. "Metalevel Reasoning." In Logic-87-8, Logic Group, Stanford University, 1987.

Gerosa M. Second Life. Melteni editore srl, Roma 2007.

Gillies D. Artificial Intelligence and Scientific Method. Oxford University Press, 1996.

Hill, P.M. and Lloyd, J.W. The Gödel Programming Language. Cambridge, MA: The MIT Press, 1994.

Hofstadter D.R., Gödel, Escher. Bach: An Eternal Golden Braid. Vintage Books, 1979.

Lanzarone G.A. "Computational Meta-languages: Theory and Applications." In Language between Theory and Technology, edited by L. Cyrus, H. Feddes, F. Schumacher, and P. Steiner. 123-34. The Deutscher Universitaets-Verlag, Jan. 2003.

Luhmann N. Die Realitaet der Massenmedien. Westdeutscher Verlag GmbH, Opladen 1996.

Nelson R.J. "Artificial Intelligence, Philosophy and Existence proofs." In Machine Intelligence, edited by Hayes, Michie, and Pao. 541-53. Ellis Horwood, Ltd. and John Wiley & Sons, 1982.

Perlis D., "Languages with Self-Reference I: Foundations (or: we can have everything in first-order logic!)." Artificial Intelligence 25 (1985): 301-22.

Smith B.C. Doctoral dissertation at MIT in 1981; published as: Reflection and Semantics in LISP. In: Procs. 11th ACM SIGACT-SIGPLAN symposium on Principles of Programming Languages, Salt Lake City, Utah, United States, Jan., 1984, pp. 23-35.

von Foerster H. Observing Systems. Seaside (CA), 1981.

Weizenbaum J. Computer Power and Human Reason. Freeman, 1976.

Weyhrauch R.W. "Prolegomena to a Theory of Mechanized Formal Reasoning." Artificial Intelligence 13 (1980): 133-70.

Winograd T. and Flores F. Understanding Computers and Cognition. Addison-Wesley, 1988.

Testing Tools of Reasoning: Mechanisms and Procedures

Bertil Rolf
Blekinge Institute of Technology, Ronneby, Sweden

How to test reasoning software?

One would expect it to be easy to show beneficial educational effects of computerized reasoning tools in courses teaching reasoning and decision making. As we shall see, the matter is complicated.

Here, I will restrict myself to software support for teaching elementary reasoning skills. There are half a dozen workable software packages of this kind, e.g., Araucaria, Athena, Belvedere, and Rationale (formerly Reason!Able).[1]

Common to such general purpose reasoning tools is a graphical interface enabling users to build a kind of tree-like graph, containing nodes representing premises and conclusions, and edges between nodes, representing logical relations. Such reasoning tools support judgment formation and decision making by externalizing and visualizing inner mental processes, enabling stepwise, openly inspectable procedures.

Prima facie, we would expect that such software would be beneficial for enhancing students' logical capacity. The software packages typically encode the graphical tools used in various analytic treatises and textbooks. They put teachers' explanatory tools in the hands of students. If the graphs used by textbook authors and argument teachers have any educational effects, we would expect even greater educational effects if those concepts in software form are put in the hands of student users.

However, in the debate about effects with and without reasoning software, conflicting or problematic claims are made. Cheikes et al. test Toulmin-based argumentation structures without software and find effect claims problematic.[2] Van Gelder and Hitchcock claim to have found software effects.[3] Baker et al. have found none.[4] Suthers attempts to link particular software features to particular effects.[5] Braak et al. have analyzed effect claims related to four different software packages. They reject the studies as inconclusive for lack of experimental rigor.[6]

This paper claims that there are two kinds of empirical testing: intercontextual and intracontextual, depending on the intended scope of inductive generalization. Intercontextual tests would need to correct for mechanisms contributing to effects. Such mechanisms are today largely unknown. Intracontextual testing is far safer. But are not its lessons tied to a specific context? It will be shown that, to some extent, its lessons generalize beyond its context.

Intercontextual and intracontextual tests of reasoning software

There are two kinds of inductive testing strategy. One kind of testing strategy aims for intercontextual conclusions, i.e., extrapolations of patterns from one educational context to others. Another testing strategy aims for intracontextual conclusions, i.e., conclusions about which effects are produced by which mechanisms in a fixed educational context or contexts similar to it.

In intracontextual testing and generalization, the idea is to fixate a package of educational mechanisms pertaining to the educational frame, the type of students, their level, program, and class, the teaching method, teacher-student communications, and student's tasks. Normally, this package of mechanisms is not explicitly identified or disentangled. The inductive test and inference is valid in relation to the package of educational mechanisms of the context at hand, but perhaps not beyond that.

The two types of induction can be used as complements. For instance, intercontextual induction may assure the teacher that, in principle, the software should work in that course. But s/he would still need intracontextual induction to accommodate the use of software so as to maximize the effects of the package of mechanisms. For most teachers, intracontextual induction can help solve educational problems connected with a certain type of course.

Intercontextual testing – the state of the art

Let us consider the state of the art of intercontextual testing. In a meta-analysis of four attempts to study effects, van den Braak et al. evaluate the results of testing four software packages: Belvedere, Convince Me, Questmap, and Reason!Able.[7]

The test procedures are evaluated with respect to internal validity and external validity. The "internal validity" of a test is related to the possibility of the test to establish causal effects in the population tested. By the "external validity" of a test, one refers to the generalization of the effects in the tested population to populations not tested. The authors supply a number of criteria for these types of validity:

> Internal validity: at least one control group, random assignments of participants and homogenization of population.
>
> External validity: draw a random sample from a population, use real world setting and stimuli, and replicate the experiment.

These criteria then form the basis of the evaluation of the tests of the software packages.

There are problems about these criteria. First, one cannot, of course, draw a random sample from a population, part of which is in the future. Second, the authors note that present unreliable measurements destroy validity of experimental results. Without reliable measurement of argumentative skills, none of the four tested packages can claim validity.

According to these criteria of validity, there are no significant results about positive effects of reasoning supporting software. The authors recommend design of future tests so as to be both internally and externally valid. Objective measures for the effectiveness of tools should be formulated. Finally, a stepwise research plan is proposed.

While it is easy to agree with the authors about the absence of conclusive effects, their diagnosis of the situation is too optimistic.

Intercontextual inductive inference depends on constancy of mechanisms

Inductively strong inferences are collections of non-deductive inferences that confer probability or likelihood on their conclusions, given that the premises in the inference are true.[8] When we use induction, we wish to draw conclusions about a population whose members had no chance of being sampled, e.g., future users of medicines or educational tools.

Induction is sensitive to causal mechanisms. Typically, we are interested in the question whether our reasoning software has larger beneficial effects on reasoning than our textbooks have. Software and textbooks take effect by partly different mechanisms that we might represent as follows:

> Educational effects of software use: S, M1, M2…Mn, S+M1, S+M2,… S+Mn.
>
> Educational effects of textbook use: T, Mk, Mk+1…Mw, T+Mk, T+Mk+1,…T+Mn.

Here, we assume that the Ms stand for mechanisms not involving software and S+Mi stands for a mechanism involving both software use and other mechanisms. Normally, it is not software as a whole that produces effects but the use of certain functions in the software packages. So we should not symbolize software effects as a unit but as a vector: (S1……Sj) with different Si representing software functions. The same holds for textbooks. We can here ignore this complexity.

Mechanisms of software use and textbook use can blend, interact, or partially overlap. For instance, there can be interaction effects, S+Mi versus T+Mj, where effects cannot be traced to Mi or Mj alone. Probably software effects are sensitive to task factors and contribute more to tasks involving complex reasoning with many premises and several layers of argument. Furthermore, reasoning software can employ educational mechanisms that do not exist or cannot be used without software. Software has many functions that textbooks do not. Many of them are technically trivial—such as undo, redo, cut, copy, paste, and save. A standardized workspace facilitates student-student and student-teacher communication in problem solving.

All inductive inference relies on constancy of mechanisms producing the outcome. When we extract properties of one population P1 and generalize them to another population P2, we need to assume that the mechanisms generating effects in P1 occur in the same proportion or exercise the same strength in P2. For instance, say that P1—unknown to us—contains 90 percent bacterial infections and 10 percent viral infections, whereas P2 contains 10 percent bacterial infections and 90 percent viral infections. Suppose that we randomize patients between experiment and control group when testing treatment of penicillin in P1 and find a significant difference between experiment group and control group. Even so, we should not expect the same difference between those exposed and those unexposed to penicillin to occur in P2.

Inductive generalizations about effects become invalid when effects are produced by partly different mechanisms. Interference of unknown mechanisms that the experimenter did not know and had no chance of controlling may invalidate inductive inferences.

Induction is critical in educational matters. There, we often have few clues about which important mechanisms there are and how they, in general, contribute to which effects.

Intercontextual testing of tools assumes another kind of causal modeling

The design and selection of a test procedure for intercontextual effects needs to be guided by causal modeling. In testing, causal mechanisms need to be specified and controlled for. Test procedures first systematized by R.A. Fisher relied on a causal modeling of interventions in agriculture. Fisher's kind of modeling was proper to agriculture, where we know little about the causal mechanisms producing effects.[9] The mechanisms controlled for were robust and relatively well known, e.g., slope or moisture. Generalizations and their limitations are fairly clear, e.g., experiments on English soil, climate, and microorganisms cannot be inductively generalized to Mediterranean soil, climate, and microorganisms or, in general, where outcome is determined by mechanisms not controlled for in the experiments.

Can the causal models used to test effects of manure in the agriculture also be used to test the effectiveness of tools used in education? I doubt this. Models of cause and effect are more precarious in the cultural and social realm than in agriculture. Causes and effects may not always generalize across contexts.

There are causal differences between adding manure to a field, controlling for slope, sun, or moisture, and having students use software to solve certain tasks. Manure, slope, sun, or moisture are natural kinds, i.e., classes where the effects are determined by natural law. The yield per acre of a field of wheat is the outcome of natural forces measured on a ratio scale.

Educational effects are conventionally delimited and produced by artifacts. Reasoning skills are evaluated according to human convention, somewhat varying from one theorist to another, from one teacher to another, or from one professional context to another. We may assume that there are family resemblances between various actors showing reasoning skills, but the classification or ranking of skills is largely a social and cultural artifact.

The causes for producing educational effects are also cultural and social artifacts. In one kind of cultural and social setting, a certain type of teacher-student interaction can produce effects that differ from those that would be produced in another setting. If the students expect and accommodate to one type of tasks and teacher instructions, they will respond in a certain way. But if not, not.

Furthermore, using tools to produce effects is different from bringing about effects by adding manure to a field. When chemicals are added to a field in order to improve on the crop, we assume that the dexterity of the agent adding them plays no causal role. We therefore rightly ascribe causal potency to chemicals, as their effects come about independent of many contextual factors. But we do not ascribe effects to tools in

themselves, for the dexterity of the user and factors relating to the aim and the working process play essential causal roles.

Moreover, we do not normally compare, say, the effects of axes versus the effects of saws, controlling for differences in wood, carpenters, aims, and work procedures. Axes can be used in indefinitely many ways, laboring with wood, and so can saws. In a few cases, their relative effects can be compared. But most often, the user purpose and user process are not quite comparable. Only exceptionally do we ascribe systematic differences in causal potency to the tools themselves.

Finally, in testing educational tools, in particular software-based tools, we are testing for what Chomsky called "competence" in distinction to "performance."[10] Competence is the storage of an abstract system of procedures or rules while performance is the output of such a system. Performance is influenced by several other factors such as memory limitations, time pressure, fatigue, lack of processing capacity. Testing for competence merely via actual reasoning performance is an indirect way of approaching the procedures underlying reasoning competence—and a blunt way at that.

Let me sum up these points, applied to education.

- The causal mechanisms connected to educational software and textbooks produce effects by means of social and cultural properties, not by means of robust, natural kinds.
- The causal mechanisms related to the educational software versus textbooks cannot be isolated from other factors, specific for the context. While some of the mechanisms are present in several contexts, the combined effect of their contribution is specific to context.
- There are too many possible educational uses or learning mechanisms related to software versus textbooks. We are interested in comparing them across a large spectrum of tasks, not merely with respect to few, standardized tasks where their relative effects can be compared.
- Education aims for reasoning competence. Competence is not stored in a black box, but we have many ways to poke into it. A teacher, designer, or tester can rely on information about what has gone into the building of competence. The output of competence can be studied in many different ways.

Given facts about induction and causal modeling, there is very little hope for intercontextual testing. Intracontextual testing is not affected by these objections. Several of the software tests actually performed can better be interpreted as intracontextual; see, for instance, some tests of Rationale/Reason!Able. When one looks more closely into effect studies, they seem to be intracontextual rather than intercontextual.[11] But intracontextual testing brings in other kinds of problems.

Does intracontextual testing make sense across contexts?

In a previous paper, I have shown how Bayesian induction enables intracontextual tests. One can infer the existence of effects from the particular way a software package is used within an educational context.[12] It may seem, however, that such tests only admit conclusions relative to that context. This would be an error. I will show why this is the case, using two examples.

First, we set as target competence the mastery of procedures enabling branch-following oral argumentation. By "branch-following," I refer to argument sequences of a certain type:

Branch-following oral argumentation	**Non-branch-following argumentation**
Proponent: Accept Thesis, because of A.	Proponent: Accept Thesis, because of A.
Opponent: Reject A Thesis because of A11 and A12.	Opponent: Reject Thesis because of B.
Proponent: Reject A11 because of A111. Reject A12 because of A121.	Proponent: Accept Thesis, because of C.

Branch-following involves pursuing arguments of the second and higher orders related to a topic before moving to a new, major topic or branch. It is desirable in order not to create cognitive overload in hearers (sometimes also speakers). Argumentation of the other, non-branch-following type is far more common in non-professional contexts. It is possible for a teacher-instructor to design instructions for oral argumentation along with reasoning tools that will improve student capacities to pursue the desired branch-following oral argumentation. Such instructions and software features may not give the desired effect the first time they are tested. Several generations of instructions and software features may need to be designed and tested.

Observe the following. The target competence consists in specific procedures of oral argumentation. They are not likely to be discovered in a standard test on critical thinking. The competence of students is theoretically described. An expert observer can identify the exercise of branch-following procedures. But a lay observer would perhaps notice only the energy of pursuit and the clarity of communication without being able to identify or diagnose which procedures are exercised.

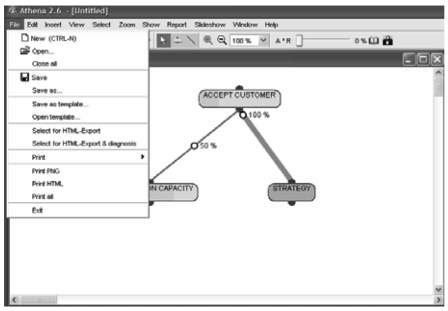
What can be generalized across contexts here? By letting students execute certain externalized procedures, documented in instructions, tasks, and software, one can help students internalize procedures of branch-following oral argumentation. Branch-following is a surprisingly complex social task, demanding metacommunication between actors mutually controlling the interchange. It is a general and informative fact of human learning capacities that complex procedures of cognitive reasoning paired with social competence can be installed via simple externalized procedures.

Another example of intracontextual design and testing is as follows. In a course where Athena was used, the teachers noted that the students did not mobilize all their knowledge to construct arguments. From their previous courses, many arguments could be drawn of relevance to the issue of debate. The teachers therefore designed a new software function and a corresponding intermediate task: to build an argument template. Students given that extra task were found to improve in breadth or argument, i.e., the desired effect was achieved. By first constructing a general template that can be saved and reopened before proceeding to specific applications of the template, students were able to mobilize previous knowledge—see Figure 1.

The target competence involved top-down procedures of inventing arguments, tied to a certain case. The target competence could be reached via interposing a template task and using template software features together with instructions.

Here, a feature of the software and a new task are combined to achieve highly specific target competence. The original software functions are extended to serve a new educational function to solve a new educational task. Of course,

Figure 1.



student learning effort and time spent in education are not identical, so a strict comparison of effects of new tools versus old would not be fair.

When we think of software and textbooks as tools to be tested, we often aim for a very specific target competence. The teacher/designer wonders whether such outcome procedures can result from the exercise of such externalized teaching and learning procedures. Experiments will make this clear. The outcome procedures and the teaching procedures normally have details far beyond the precision of standard tests for critical thinking.

The key feature of the causal models underlying such intracontextual testing is that one type of externalized procedures is used to bring forth competence, stored as procedures in human actors. Many procedures are specific to the context, relying on a multitude of educational features. But while the sentences, so to speak, are specific to the context, the words making them up are not. The causal effects are produced by interacting complexes; the elements making them up are transferable to new context and can be put to good use there. The generalized conclusion is that such tools in the hands of such craftsmen, using such techniques can bring forth new features of reasoning competence.

How to improve on intracontextual testing?

I propose that intracontextual testing can be strengthened in several ways. A first recommendation for intracontextual testing is that one sticks to a narrow and operationalizable definition of which parts of reasoning competence one would like to improve. These parts can be analyzed into a class of more elementary procedures. Next, one attempts to find a simple way to form a package of tasks, instructions, necessary background knowledge, and software features so that the execution of these procedures is facilitated. This is, in principle, a complex, causal model of which procedural mechanisms together could produce the desired storage of reasoning procedures. Finally, one tests whether the outcome when these procedures are used is an improvement towards the desired objectives relative to what is normally achieved without the package. There is an implicit reference to the course context where most of the other features are held constant.

A second recommendation is to document the complete packages used to achieve effects. At present, the documentation of testing conditions is faulty. For intercontextual testing, this is disastrous; for intracontextual testing, this is a missed opportunity of learning across contexts. The learning involves not a generalization of conclusions about effects but about which procedures can bring forth which competence.

A third recommendation is to try to isolate which desired effects on reasoning procedures can be achieved by which combinations of software features, background knowledge, tasks, instructions, and feedback. For instance, it seems likely that argument diagramming can be used to improve students' ability to analyze argument structure. But does software for argument diagramming contribute anything over paper and pencil?[13] A more finely grained approach will enable teachers to maximize desired effects by minimizing the costs of means.

Finally, in testing of software versus textbooks, one needs to be clear about which questions are interesting to ask and promising to try to answer. Intercontextual testing is neither, I believe. The questions asked depend on a peculiar view of mechanisms and functions of tools. The answers cannot be extracted and applied across contexts. The ambition of intercontextual testing is, however, laudable in its aim for generalizable knowledge. I have suggested how to design intracontextual testing in ways that make aspects of tool use interesting across institutional and educational contexts.

The answers from tests of educational effects do not generalize across contexts. Only the questions, the methods, and a stance of conscious, critical use of reasoning tools do.[14]

Endnotes

1. For an overview, see Maralee Harrell, "Review Article: Using Argument Diagramming Software in the Classroom," Teaching Philosophy 28:2 (2005): 163-77 and Cris Reed, Douglas Walton, and Fabrizio Macagno, "Argument Diagramming in Logic, Law and Artificial Intelligence," The Knowledge Engineering Review 22:01 (2007): 87-109.
2. Brant A. Cheikes et al. "An empirical evaluation of structured argumentation using the Toulmin argument formalism," Technical Report MTR 04B0000074 (Bedford, MA: MITRE, Center for Integrated Intelligence Systems, 2004). http://www.the-mitre-corporation.net/work/tech_papers/tech_papers_04/04_1032/04_1032.pdf (accessed July 23, 2007).
3. Tim van Gelder. "The efficacy of undergraduate critical thinking courses. A survey in progress 2000." http://www.philosophy.unimelb.edu.au/reason/papers/efficacy.html (accessed July 23, 2007). David Hitchcock. "The Effectiveness of Computer-assisted Instruction in Critical Thinking." http://www.humanities.mcmaster.ca/~hitchckd/effectiveness.pdf (accessed June, 2005).
4. Michael J. Baker et al. "Designing a Computer-supported Collaborative Learning Situation for Broadening and Deepening Understanding of the Space of Debate." Proceedings of the Fifth International Conference of the International Society for the Study of Argumentation (Amsterdam: Sic Sat Publications, 2002): 55-62.
5. Daniel D. Suthers. "Representational Guidance for Collaborative Learning." Artificial Intelligence in Education. 11th International Conference on Artificial Intelligence in Education, edited by Heinz U. Hoppe et al. (Amsterdam: IOS Press, 2003), 3-10.
6. Susan van den Braak et. al. "A Critical Review of Argument Visualization Tools: Do Users Become Better Reasoners?" Workshop Notes of the ECAI-2006 Workshop on Computational Models of Natural Argument (CMNA VI), edited by Floriana Grasso et al. (Riva del Garda, Italy, 2006): 67-75.
7. See previous note.
8. Brian Skyrms. Choice & Chance: An Introduction to Inductive Logic (Belmont: Wadsworth, 2000), 17.
9. Gerd Gigerenzer et al. The Empire of Chance: How Probability Changed Science and Everyday Life (Cambridge: Cambridge University Press, 1989), 3.2.
10. Noam Chomsky. Aspects of the Theory of Syntax (Cambridge: The MIT Press, 1965), §1.
11. See, for instance, Charles R. Twardy, "Argument Maps Improve Critical Thinking." Teaching Philosophy 27:2 (2004): 95-116.
12. Bertil Rolf. "Testing Reasoning Software: A Bayesian Way." Selected Papers from the European Computing and Philosophy Conference ECAP 2005 4 (2005): 328-32.
13. A question raised by Maralee Harrell, "Using Argument Diagrams to Teach Critical Thinking Skills," draft, http://www.hss.cmu.edu/philosophy/faculty-harrell.php (accessed June 2007).

14. This work has been granted support from the Swedish Environmental Protection Agency. A previous version has benefited from comments and questions by Anders Törnqvist and Marvin Croy.

Book Review

Virtually Obscene: The Case for an Uncensored Internet

Amy White. Jefferson: McFarland & Company, 2006.

Reviewed by Melissa Winkel
University of Illinois–Springfield

Censorship of Internet pornography is the subject of White's new book, Virtually Obscene: The Case for an Uncensored Internet. She examines points for and against censorship of Internet pornography and ultimately arrives at the conclusion that attempting to control any content on the Internet would cause greater harm than good.

The author provides three main arguments against censorship: pornography is not harmful; censoring Internet pornography would lead to the censorship of other forms of expression; censoring the Internet, even if it were desirable, would be difficult if not impossible to accomplish technologically. Before addressing these points White explains why her case does not rely on the free speech argument. Interestingly, despite arguing against censorship, the author demonstrates the weaknesses found in the most common argument given by those who share her view. In short, the author chose a very controversial topic to explore and she is successful at revealing the flaws in each side's arguments.

White begins by examining the most common argument against censorship, the First Amendment. Free speech, we are reminded, does not mean freedom of all speech as in the case of perjury, libel, defamation, or when used to incite unlawful violence, among others. For these reasons it is argued that free speech does not warrant special status among liberties. Further, the author argues that Mill's argument that free speech is required to uncover truth was not designed to protect pornography or other explicit materials, which can be said to have little, if any, truth value. Likewise, because most Internet pornography also has no political value, the arguments for free speech based on democracy and self-governance lend no support to the argument for unregulated Internet pornography. By the end of this section it is clear that the most cited argument for an uncensored Internet is greatly flawed.

White begins the main thread of her argument by trying to prove that pornography is not harmful; in particular, she attempts to respond to the claims that online porn is harmful to children, women, or to the moral community. Following, the Harm Principle, most famously outlined by John Stuart Mill:

> The only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not sufficient warrant.

With this principle established, it is argued that the harm to children, women, and the moral environment arguments are not sufficient to warrant the censorship of Internet pornography because the pornography does not cause **direct** harm. As such, White addresses arguments that people may be indirectly harmed by pornography—children who may be accidentally exposed to it; women who are degraded by it; society in general that suffers moral decay as a result of wide availability of sexually explicit content.

While addressing the harm to children argument, the author acknowledges that a vast number of people argue the Internet should be regulated because Internet pornography is harmful to children. After noting that pornographic and/or obscene material may not in fact be harmful at all, she asserts risks to such exposure can occur anywhere, from cyberspace to the local mall to one's own backyard. Further, she notes that Internet access is voluntary and parents wanting to protect their children from such materials can install filters on their computers or not have Internet access in their homes, both of which are viable options. While her assertion that pornography and/or obscene materials may not be harmful at all is controversial, the fact remains that a great deal that can be done to prevent children from being exposed. Thus, the author's conclusion that censorship based on harm to children appears to be an extreme solution when there are relatively easy ones readily available is of much interest. It is worth discussing further how parents can oversee their children's Internet exploration rather than the government managing the Internet for all citizens.

The harm to women argument is more complex. Three arguments are presented by White against the theory that Internet pornography should be regulated because it is harmful to women: pornography production conditions are the same as in other job markets; pornography does not lead to sexual violence; and pornography treats women as equals and thus does not degrade women. By attempting to demonstrate there is no proven link between pornography and violence against women, the author severely weakens her opponent's argument that pornography directly causes such harm. To further weaken the argument, both feminists and pornography actresses are cited as evidence pornography is a form of liberation and an autonomous choice. I believe, however, one important point is missed—harm to women can and does occur in the pornography industry. Information on such abuses is readily available. For example, pornographic producer and director Janet Romano, in an interview with PBS, openly admits to physically and verbally abusing other female actresses and goes on to state, "I used to be exploited when I did movies. So if someone's going to do it, I might as well" (American Porn 1). In addition, a Florida man was recently convicted of kidnapping and rape after forcing his estranged wife into the woods and raping her. His intent was to sell video of the ordeal to pay off his debts. Yet White fails to ever acknowledge this fact by citing one case or providing one example. Rather than providing direct, factual evidence of such abuses and refuting them, the author attempts to discredit this fact by using statements such as "it is claimed," "as 'evidence' of such abuse," and "as for the claim…" (White 92). Even if such cases are not commonplace, they should have been presented and addressed.

The author swiftly responds to the claim that pornography causes harm to the moral environment by demonstrating that "community standards" cannot be used to determine what is or is not harmful in a given society. The pluralistic nature of geographic communities and the incredible diversity that exists on the Internet creates a multitude of differing moral views and, as such, it is unlikely any "community standard" would be agreed upon. Hence, community standards may not serve as the basis for Internet censorship.

The author's second argument, that censoring pornography would result in more harm than good, is by far the strongest.

By inadvertently censoring non-pornographic but explicit material, many famous and valuable works of art that contain nudity or homosexuality could be removed from the Internet by imperfect government filtering systems. Further, following Mill's description of freedom of expression, the author demonstrates the liberty of many people would be hindered by censorship because doing so would limit their autonomy. Also presented is the argument from John Locke that one has ownership of one's body. Based on this reasoning, it is argued that censoring the Internet could infringe on a property right. The author also argues that the biased manner in which censorship would be regulated—by a majority of mature, Caucasian men in the legislature—would lead to the suppression of minority or controversial views such as information on homosexuality, birth control, or abortion. The author is also careful to acknowledge that liberties cannot and do not always carry the trump card; they are only to be tolerated and pursued so long as they do not cause direct harm to others, which is the case with Internet pornography as White shows in her first argument.

Finally, the third argument that the technological aspects of censoring the Internet are complex is valid, but it does not render censorship impossible as seems to be asserted. While it is true that producers of pornographic material could move offshore and the task of monitoring what millions of people are viewing would be daunting, there are possible solutions. For example, it would not be difficult to limit people's ability to pay for pornography online. In fact, this is how the government severely hampered people's ability to participate in online gambling.

Overall, White's book takes a close look at a hotly debated topic. Most arguments for censorship are well-presented, thoroughly investigated, and easily shown to have weaknesses. However, others, such as the harm against women argument, could have been presented a bit more carefully. Further, the slippery slope argument is dismissed as faulty by the author when used in favor of a position she dislikes but then used to support the position she favors. More specifically, in the final chapter the author states that "[t]o argue that the potential of the Internet need not be hindered by regulation is also to ignore the fact that one of the potentially most beneficial aspects of the Internet is its ability to escape governmental control," and "if the Internet is made to yield to regulation, cases like this may no longer be commonplace" (White 141-42). While this is not as direct as advocates of the negative slippery slope argument, the author is implicitly making the same assertion; if X then Y. Yet there is no evidence this will occur. In fact, she presents contrary evidence when describing China, where the Internet is regulated (White 21), as a place where minorities often use the Internet to speak out against the government (White 141). The subject of censorship on the Internet is important and should be investigated. While this book is built around a convincing and powerful line of reasoning, it succeeds only in part.

Acknowledgments

I would like to thank Peter Boltuć, Faisal Nsour, Linda Williams, and Dmitir Pekker for their valuable insight and thoughtful contributions to this review.

References

"American Porn." Frontline. PBS. 2 Feb. 2002. Transcript available at: http://www.pbs.org/wgbh/pages/frontline/shows/porn/etc/script.html

Raynor, Jessica. "Jurors Say Porn Video Was Rape." Florida Today. 13 Jul. 2007. Local News. 20 Jul. 2007. http://www.floridatoday.com/apps/pbcs.dll/article?AID=/20070713/NEWS01/707130343

White, Amy. Virtually Obscene: The Case for an Uncensored Internet. Jefferson: McFarland & Company, 2006.

Response to Melissa Winkel

Amy White
Ohio University

While I thank Melissa Winkel for her review of my recent book Virtually Obscene: The Case for an Uncensored Internet, I believe a few points deserve comment. Winkel expresses concerns that I do not address the fact that women are sometimes harmed in the pornography industry. In Virtually Obscene, I acknowledge that there may be cases where women have been harmed in the pornography industry. I also suggest that such harm may be reason to monitor the working conditions in the industry. The pornography industry should be subject to the same scrutiny that is present in other workplaces. However, there are pornography production companies that offer decent environments, and harms are less commonplace in pornography than in many other work environments.

Winkel clearly misunderstands the nature of a slippery-slope argument. While it is true that I discuss the ability of some users to escape this governmental control, it doesn't create a contradiction in my argument, as Winkel seems to suggest. I simply argue that minority or unpopular views would no longer be commonplace and easy to find if the Internet were regulated, not that some would not escape regulation. The history of the Internet is filled with cases of users bypassing controls; however, most users are not technologically knowledgeable enough to circumvent regulatory controls.

NOTES

InPhO: The Indiana Philosophy Ontology <http://inpho.cogs.indiana.edu/>

Cameron Buckner, Mathias Niepert, and Colin Allen
Indiana University–Bloomington

Introduction

The goals of the Indiana Philosophy Ontology (InPhO) project are to build and maintain a "dynamic ontology" for the discipline of philosophy, and to deploy this ontology in a variety of digital philosophy applications. Automated information-retrieval methods are combined with human feedback to build and manage a machine-readable representation (i.e., a "formal ontology") of the relations among philosophical ideas and thinkers. The applications we hope to develop that will employ the ontology include automatic generation of cross-references for Stanford Encyclopedia of Philosophy (SEP) articles, semantic search of the SEP and other philosophical resources (including guided searching with Noesis), conceptual navigation through the SEP using information visualization techniques, and web access to the biographical and citational information contained in the InPhO. Moreover, we will archive the dynamically generated versions of the ontology, so we can digitally and dynamically track changes to the discipline of philosophy over time.

By focusing our initial efforts on the SEP, we build upon the most-developed and highest impact project in Digital Philosophy. Yet the full digital potential of the SEP is far from being realized. It is a "dynamic reference work" with 900

published articles comprising over 9.5 million words, growing at approximately 100,000 words per month. The SEP is already beyond the point where it can be comprehended easily (if at all) by any single individual. As it continues to grow, new ways to organize, visualize, navigate, and search the rich discipline-specific content are needed. The SEP's unique combination of scale, authoritativeness, and open access makes it an ideal starting point for developing an ontology for philosophy.

The InPhO contains four sub-ontologies: Thinker, Idea, Document, and Organization. The Thinker sub-ontology categorizes persons of interest to philosophy along professional lines and captures a wide array of biographical information in a formal knowledge base. The Document sub-ontology will consist of a bibliographical database capturing information about philosophical publications, and the Organization database will record organizations (such as universities and societies) of interest to the domain of philosophy. The most important sub-ontology, however, is the Idea sub-ontology, which categorizes keywords (and phrases) corresponding to philosophical ideas along sub-disciplinary lines. For example, philosophy decomposes into: ethics, logic, metaphysics, philosophy of mind, and so forth. Each subdiscipline in turn divides into a number of issues considered fundamental to work in that area; for example, philosophy of mind is currently divided into: mental content, metaphysics of mind, consciousness, philosophy of psychology, and philosophy of artificial intelligence. Each of these subdivisions in turn divides into issues considered fundamental for work in that area, and the process is repeated until a level of specificity is reached that is sufficient to categorize the most specific keywords of the SEP. The ontology is intended to be dynamic, with additional divisions being semi-automatically introduced as necessary to accommodate the additional content provided by new and revised SEP entries.

Implementation

Many projects in the digital humanities face problems that are associated with managing large bodies of sophisticated text. Digital philosophy projects present special problems because they contain abstract language that is especially difficult to analyze using automated methods, and because the structure of the discipline itself is often a matter of dispute. Several approaches to the general problem of classifying texts have been proposed. The approaches can perhaps be divided into two classes: full automation and harnessing social collaboration. Fully automatic analysis of text is a hard problem in artificial intelligence. The most popular current approaches rely on statistical co-occurrence of keywords, but these methods fail to reach the levels of accuracy required by the standards of academic scholarship. Instead of automation, approaches that rely on social collaboration have been more popular recently. These approaches create and harness metadata (machine-readable descriptions of the primary content) by soliciting the collaboration of the users of online encyclopedias and databases. This category includes the well-known wiki-based approaches (such as Wikipedia) and social tagging methods used to create "folksonomies" (such as with del.icio.us). These approaches, however, also fail to produce results that meet the standards of scholarly review—as wiki-based approaches face the persistent problems of inaccuracy and vandalism, and social-tagging approaches also result in heterogeneous taxonomies that are usually unsuitable for automated reasoning.

Our proposed solution is to use automated statistical methods to generate metadata "hypotheses," which are then reviewed by domain experts. Specifically, for applications involving the SEP, this means review by the philosophers who serve as the SEP's authors and editors. We have devised algorithms that estimate the semantic similarity and relative generality of any two keywords in the SEP based on their patterns of co-occurrence in SEP entries. These statistics can be used to estimate the relative taxonomic relationship between any two keywords in the encyclopedia. These taxonomic hypotheses will then be confirmed or falsified by authors and editors through feedback forms that will be integrated with the SEP's standard document submission and review process. This feedback is stored as statements in first-order predicate logic and then aggregated from many experts into a collective knowledge base. The knowledge base is then passed to an automatic reasoning system that uses the aggregated feedback to determine the optimal location to classify keywords in a predetermined taxonomic scheme. (For technical details see Niepert et al. 2007[1]; available online at http://inpho.cogs.indiana.edu/.)

In the present version of our system, the reasoner only classifies keywords in the pre-determined taxonomy, and the taxonomic scheme itself is coded "manually" from sources that include contributions by SEP editors and several excellent, web-based annotated bibliographies also maintained by SEP editors. In future versions, however, we will explore using automatic methods dynamically to infer the taxonomic scheme itself (either in part or in whole). An initial step in this direction is to use statistical measures to generate recommendations for subdividing categories that have become too large or heterogeneous. These recommendations would be reviewed by the appropriate domain experts.

The distinguishing feature of this approach is its emphasis on the SEP's most valuable informational resource: the domain experts that serve as its authors and editors. At every stage, information must be approved by experts before going "live" into the SEP system. A central challenge to our project has been to get the most informational gain out of interaction with experts without placing undue demands on their time—it is the challenge of efficiently asking the right questions and making the most effective use of the answers.

Request for Feedback

Although we have focused initially on the needs and strengths of the SEP, we intend to publish InPhO-based applications that will serve a wide variety of philosophical needs and interests. Because the data to be assimilated are much broader than the SEP itself, and because we cannot tax the SEP's authors and editors to provide feedback on everything that might be included in the complete InPhO, we will also be soliciting feedback from all interested philosophers, who will be asked to evaluate the deliverances of the automated methods. Information provided in this way can also be exploited for applications involving the SEP, but in such applications it will be marked as "provisional" until reviewed by a known expert among the SEP contributors. Given the high quality pool of potential evaluators that exists among APA members, however, we expect this information to be useful in its own right, and ultimately to simplify the task of the SEP's contributors. Another way in which we will invite contributions from philosophers at large will be in editing and submitting bibliographic items for the Document sub-ontology using a system that we expect to announce in 2008.

Finally, by publishing regular updates of the InPhO in a standard XML format using the Web Ontology Language (OWL), we hope to inspire others to develop new applications of the ontology, and to develop alternative taxonomic schemes. We believe that alternative schemes may reasonably coexist with the InPhO as there is more than one way to organize the discipline. Our view is that one ontology is better than none, but if our particular way of taxonomizing the discipline is controversial, so be it! Let it be a challenge to others to improve

upon our efforts, for we believe it is only through iterative development and competition that philosophers will get the digital tools they need.

Endnotes

1. M. Niepert, C. Buckner, and C. Allen. "A Dynamic Ontology for a Dynamic Reference Work." In Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, JCDL 2007, edited by E.M. Rasmussen, R.R. Larson, E. Toms, and S. Sugimoto. Vancouver, British Columbia, 288-97.

The Pathways School of Philosophy

Geoffrey Klempner
International Society for Philosophers

In the summer of 1995, being an unemployed philosopher with nothing better to do, I decided to start my own school of philosophy.

I'd heard on the academic grapevine that distance learning was the coming thing, but in my plans the Internet hardly figured. This was to be an old-fashioned correspondence course with students receiving course units in the post and sending off assignments to be marked.

The previous year, my book Naive Metaphysics[1] had appeared with hardly a sound, despite an enthusiastic endorsement from my erstwhile mentor David Hamlyn.[2] A month before, I'd finished teaching a metaphysics course for final year undergraduates as a guest lecturer at Sheffield University, but there was no permanent post in the offing. I had little inclination to try another shot at publishing. The idea of writing for a captive audience appealed to me.

Many years before, when I was a graduate student at Oxford University, I once remarked half-jokingly to my D.Phil thesis supervisor John McDowell that I would love to have a school of philosophy "like Plato or Aristotle." McDowell agreed that—knowing me—that was probably the only place I would be happy. That exchange has always stuck in my mind as a turning point in my academic career.

In a nutshell, the aim of Naive Metaphysics is to demonstrate the truth, or rather half-truth, in solipsism. Rather ironic, considering the life I have lived since then. One of the more popular essay questions from the Pathways Introduction to Philosophy Program, Possible World Machine[3] is, "How do you know that the author of these words has a mind?" Only a handful of my students have had the chance to meet the author of those words.

In order to proceed with my plan I first needed to do some market research. A circular sent to philosophers at all the university departments in the UK brought some encouraging feedback. I put together an information pack with a few choice quotes and placed a postage stamp sized advertisement in the London Sunday Times: "Pathways to Philosophy—an exciting new development in distance learning."[4]

Out of thirty replies, which I received over the following week, three plucky students enrolled. Only the first unit from each of the six planned fifteen-unit courses had actually been written, but I was confident in my ability to keep up the supply of course units in response to demand; I've never suffered from writer's block.

I spent the next two years churning out course units, in between responding to student notes and essays. During that time I learned to love Apple Macintosh computers. I didn't yet have a computer of my own, so I spent long days and evenings in the Sheffield University Computer Centre.

That was also the time I discovered the Internet.

It took a while to put two and two together. By August 1997, most of the planned ninety course units had been written—around half a million words. Sheffield University kindly lent me some web space and I built a web site, the "Pathways to Philosophy Distance Learning Programme," with help from the Sheffield Computing Service's four page "Guide to HTML for Beginners." The six Pathways remained unchanged; the only difference was in the method of delivery. The course units and essay questions were reviewed by then Sheffield Professors Peter Carruthers and David Bell.

Since then, Pathways has introduced two new study tracks: an Associate and Fellowship for self-devised programs of study,[5] and support for the Diploma and BA (Hons) in Philosophy offered by the University of London External Programme.[6]

Like the six Pathways, the Associate and Fellowship are not university accredited, but instead validated by the Board of the International Society for Philosophers,[7] a society which I launched in March 2002 with the help of academic friends and supporters of Pathways. Successful essay portfolios submitted for the Associate and dissertations submitted for the Fellowship are archived on the Pathways web site.

To give some idea of the relative length of the University of London Diploma and BA courses, one Pathways program or Associate portfolio is roughly equivalent—in terms of study time required—to one UoL Diploma or BA module. The Diploma consists of four modules, while the BA consists of ten. A hard-working distance learning student can expect to complete two UoL modules—or two Pathways—in a year.

One of the key features of all three Pathways study tracks is that our students receive an 800-word letter from their mentor in response to each assignment: for example, notes on a course unit or an essay. At the present time, I am responsible for half the total teaching load, the rest is done by volunteer graduate students who teach on the six Pathways in return for my supervision of their work towards the ISFP Fellowship.

To date, students have joined Pathways from over sixty countries, mainly thanks to the high profile of the Pathways web sites.

Over the twelve years that Pathways has been running, I have learned quite a bit about what distance learning students are looking for—at least in a philosophy course. The majority seem reluctant to get involved in online forums or conferences. What they value most is the opportunity for one-to-one dialogue (and some of them are damn good at it too).

There is an interesting dynamic that becomes apparent if you look closely at the way people behave in one-to-one email correspondence compared with online forums. In email correspondence, one exercises tact and restraint. It takes time to get to know someone when the only input is words on a screen. By contrast—and to the despair of many forum moderators—people in forums love to sound off. And for the very same reason: you have no face, no bodily presence. The only consequence of irresponsible behavior is more words on a screen, which you can switch off at will. This gives the participant a false sense of invulnerability. On some forums that I have visited, the lack of respect is palpable.

Perhaps I have a distorted view because the Pathways web site acts as a filter for potential applicants. We have our own online conference, but it is given a low profile and run on a strictly voluntary basis with no credits for "successful" participation (whatever that means).

This totally contradicts the current accepted wisdom in distance learning. All the talk nowadays is of the great opportunities offered by the latest conferencing and interactive software. By contrast, all a wired Pathways student needs is an

email address. I'm not knocking the alternative, but I still wait to be convinced.

In 2001, I had the opportunity to explain the Pathways approach at the European Education Technology Forum organized by University College Dublin. In my handout I wrote:

> Pathways was created as a solution to a problem: how can one work in philosophy?
>
> I had no interest in writing for an audience of academic philosophers. Yet I realized I needed an audience for my work. Pathways was launched as a quest to find that audience.
>
> Pathways is unique for several reasons.
>
> It is a world class distance learning program which has arisen outside university structures. The majority of students who enroll for Pathways have no special desire to gain a qualification, but do so purely for the love of the subject. Many are already highly qualified in other fields.
>
> Pathways was conceived as a one-to-one dialogue between student and mentor, following the Socratic ideal. The form of the program is thus determined by the unique character of philosophy itself.
>
> Pathways tuition is designed to be labour intensive, at a time when universities have been looking to distance learning and computer technology as a way of increasing the throughput of students per lecturer hour. Yet Pathways is entirely self-financing, receiving no grant aid of any kind.
>
> Pathways is run as a business. It has to pay its way. In case of failure, there is no safety net. It would be interesting to see what would happen if professors faced dismissal if they failed to make a profit!

My deliberate intention was to be provocative. In the 1999 introduction to my weblog "The Glass House Philosopher"[8] I described myself as an "Internet sophist." When I repeated this to the other participating philosopher in the hotel bar the evening before the conference he replied curtly, "Well, I'm an anti-sophist!"

That was the end of that conversation.

My presentation consisted mainly of a tour of the Pathways web sites, explaining how the Pathways idea developed—the e-journal Philosophy Pathways, "Letters to My students," "Pathways How-to-do-it Guide," "Pathways Essay Archive," and, last but not least, "Ask a Philosopher," originally launched in 1999, which, though staffed mainly by graduate students, manages to give the Amherst guys[9] a run for their money.

I also tried to explain my motivation. What was in it for me? I had (and still have) no great ambitions to publish. I just wanted to do philosophy the best way I knew, by writing letters—following the example of my philosophical heroes:

> Since the Middle Ages and before, philosophers had produced masses, volumes of letters. Some of the most precious documents we possess about the modern philosophers such as Descartes and Leibniz are the letters they wrote. To all and sundry. People who were asking them about their philosophy. Students they took on, or people who were working in other fields.
>
> And I had this...crazy idea that when I wrote to my students—incidentally, writing to students isn't anything like what you imagine in a course. When a student sent me a piece of work I would write an 800-1,000 word letter in reply, and in the beginning I was taking up to three hours to do it. Multiply that by 20 and that's just one student!—I had this idea that if at some future date someone was going to collect my works, I wouldn't be embarrassed to see the letter, amongst those works. So that every letter that I wrote was an attempt to do philosophy in as honest a way as I could.[10]

After my presentation, one member of the audience remarked dryly that the Pathways model would be difficult to implement in a university department because of the teaching load. Heads nodded and there was a ripple of polite laughter.

My reply was simple and to the point: "Get your students to do the teaching!" I didn't just mean the graduate students, but second and final-year undergraduates. They could only benefit from the experience, I said. From the audience reaction, I could see that this was obviously a novel idea.

At this point, the academic reader is probably grimacing at the thought of university departments taking advantage of the knowledge and teaching abilities of the average undergraduate student. Are undergraduates going to grade assignments and mark exam papers? Where would that lead?

Where indeed.

Maybe this is an idea whose time has not yet come. I would argue that the current widespread student cynicism and apathy, and the growing service industry of cheating and essay writing sites, is largely a consequence of the misplaced emphasis on getting the right letters after your name. Fierce competition for places in the best graduate schools results in too much importance being placed on the process of weighing and measuring the individual student's academic performance, and not enough on the aspects that cannot be measured—the sheer joy of learning and enlarging one's mind.

Pathways students are different. We have doctors, lawyers, priests and rabbis, school teachers, programmers and IT consultants, business and marketing executives—as well as a handful of university professors. As one would expect, there is a noticeable bulge around the forty-somethings, but ages range from fifteen to the mid eighties—all seeking refreshment at the ancient well of philosophy. It is an incredible joy and privilege to have the opportunity to engage these people in dialogue.

At the beginning of 2006, Pathways moved from the Sheffield University web site to commercial web hosting. Just last week, I decided to break the last remnants of the umbilical cord and changed my email address from sheffield.ac.uk to Fastmail. I do get annoyed when people assume that Pathways is run under the supervision of the Sheffield Philosophy Department, even though I am proud to have worked there.

My career has recently taken a turn in the direction of the philosophy of business and business ethics, following the launch in 2002 of a second Pathways e-journal, Philosophy for Business. It's still too early to tell whether the new graft will take, although I've enjoyed my business trips. The most recent was in March for a presentation at a one-day conference in Prague organized by the British Chamber of Commerce Czech Republic on the topic of "Social Responsibility for Small and Medium Sized Enterprises." Corporations have a lot more money to spend, but I find business people hard to fathom. Perhaps only time will tell how much of a sophist I really am.

Endnotes

1. Naive Metaphysics: a theory of subjective and objective worlds. Avebury Series in Philosophy 1994. Now available as a PDF download from http://www.philosophypathways.com/download.html
2. Avebury Flyer http://www.philosophypathways.com/programs/book3.html
3. Six Pathways http://www.philosophypathways.com/programs/pak2.html
4. "Pathways to Philosophy: Seven Years On." Practical Philosophy Journal of the Society for Philosophy in Practice 6:1 (April 2003). Online at http://klempner.freeshell.org/articles/pathways.html
5. Associate and Fellowship Awards http://www.philosophypathways.com/programs/soc.html
6. University of London Diploma and BA via Pathways http://www.philosophypathways.com/programs/lond.html
7. Board of the International Society for Philosophers http://www.isfp.co.uk/international_society_4.html
8. Glass House Philosopher http://www.philosophypathways.com/glasshouse/ (currently "in quarantine awaiting review and reconstruction")
9. AskPhilosophers http://www.askphilosophers.org
10. 2001 European Educational Technology Forum, Video Highlights.

Pathways web pages

Pathways School of Philosophy
http://www.philosophypathways.com
International Society for Philosophers
http://www.isfp.co.uk
Philosophy Pathways e-journal
http://www.philosophypathways.com/newsletter/
Letters to My Students
http://www.philosophypathways.com/letters/
How-to-do-it Guide
http://www.philosophypathways.com/guide/
Pathways Essay Archive
http://www.philosophypathways.com/essays/
Ask a Philosopher
http://www.philosophypathways.com/questions/
Philosophy for Business e-journal
http://www.isfp.co.uk/businesspathways/

NA-CAP@Loyola 2007

Matt Butcher
Loyola University Chicago

This year's annual North American Computers and Philosophy (NA-CAP) conference was held at Loyola University Chicago on July 26-28. It was hosted jointly by Loyola's philosophy and computer science departments.

NA-CAP is the North American chapter of the International Association for Computers and Philosophy (IA-CAP), an international body interested in promoting philosophical examination of computing and related fields. IA-CAP's commitment to scholarship is most evident in their hosting of three annual conferences—one in North America, one in Europe, and one in the Asia/Pacific region.

The themes for this year's NA-CAP conference were the Free and Open Source Software (FOSS) and the Open Access (OA) movement in the realm of scholarly journals.

The FOSS movement is a recent trend in the software world where computer programmers release not only binary executable programs, but also the source code (the blueprints) for the software. It's not just the source code, however, that makes a program Free or Open Source. The software must also be released (or given away) under terms that allow anyone to use the software, modify the source code, build derivative works, and redistribute the software (modified or unmodified). Advocates of this method of distribution cite many reasons for using such a model, some of which are ethical (we ought to grant individuals freedom to use and manipulate software), and some of which are practical (such a process results in better software with fewer bugs).

Similarly, the Open Access movement proposes that journal articles—a market made distinctive by the fact that it is not royalty-driven—ought to be released under special conditions whereby authors (and journal publishers, and other online information sources) allow other scholars free online access to their articles. Such a system, argue OA's proponents, fosters research, increases the visibility of the article (and the author), and reduces expenses. All of this can be achieved without compromising the integrity of juried journals, the publishing process, or the scholarly environment.

Over forty papers were presented this year. Special panels on software development ethics, Wikipedia, and the ethics of FOSS/OA highlighted this year's theme. While there were many papers focusing on various philosophical aspects of OA and FOSS, the topics that have long served as the mainstays of NA-CAP scholarship were also represented. Panels on ethics and computers, metaphysics, artificial intelligence, robotics, scholarly online resources, and electronic teaching resources were vibrant exhibitions of recent philosophical work in these areas.

Also of interest, during the conference an exploratory committee was formed for the purpose of investigating the possibility of instituting an international graduate certificate in one of IA-CAP's areas of interest, such as in philosophy and computers. The committee, composed of Luciano Floridi, Marvin Croy, Ron Barnette, Peter Boltuć, Gordana Dodig-Crnkovic, Gaetano Aurelio Lanzarone, Keith Miller, and Vincent C. Müller, discussed the theory, tools, and technologies that could possibly drive such an effort.

This year's keynote speakers were Peter Suber (SPARC/Earlham College), a philosopher and well-respected advocate of the Open Access movement in scholarly journals, and Richard Stallman, the founder of the Free Software Foundation, and the originator of the Free Software Movement. Rory Smead (University of California–Irvine) was awarded this year's Goldberg Award for his paper "The Evolution of Cooperation in the Centipede Game with Finite Populations." Anthony Beavers (Evansville) acted as the program chair, with Thomas Wren, Matt Butcher, Konstantin Laufer, and George Thiruvathukal (all from Loyola) as local hosts for the conference. Marvin Croy (University of North Carolina–Charlotte) coordinated many of the conference details. Video coverage of the conference can be found at http://na-cap.osi.luc.edu.