Peter Boltuć
University of Illinois–Springfield

A few months ago I became preoccupied with the question of what the ontological status of web-based objects is. I started asking around and, as a result, we are lucky enough to publish two very different articles that tackle this issue, though indirectly. Lynne Rudder Baker’s paper on “The Shrinking Difference Between Artifacts and Natural Objects” could serve as an example of clear philosophical writing. Baker argues that there is no ontological gap between “natural” and “artificial” or “mind-dependent” and “mind-independent” substances; she also shows how this ontological argument is relevant for the kind of artificial objects the readers of this Newsletter are particularly interested in. Harry Halpin’s paper on “Philosophical Engineering: Towards a Philosophy of the Web” is of a distinctly different type: more of a broad synthesis than Rudder Baker’s tightly argued contribution, it unabashedly tackles the major philosophical issues about the Web by drawing connections to other fields, such as embodied and classical artificial intelligence. This paper, well grounded in the work of A. Clark, B.C. Smith, M. Wheeler, T. Berners-Lee, and other pioneering authors, is an overview of this new and fascinating field, and as far as I can tell at this point, it looks like a roadmap or blueprint of a groundbreaking project.

Next, we present the study by Jean Gabriel Ganascia, “In silico Experiments: Towards a Computerized Epistemology.” This article follows up on the discussion of issues of naturalized epistemology opened in this Newsletter by Dodig-Crnkovic’s article (vol. 06, no. 2). Ganascia analyzes the role of in silico experiments conducted in biology, which “reverse the traditional conception of experimentation.” He compares them to the other kinds of thought experiments or simulations and shows that they encompass both the “informational approach” and the “computational approach.” The author shows persuasively that the prevalence of computers in our world has clear implications for epistemology. Thus, the articles section starts with heavy metal ontology of the clearest sort and ends gently with a form of historically grounded naturalized epistemology.

The second section of the Newsletter is devoted to discussion of the lead papers from the last two issues. In her discussion on G. Harman’s “Explaining an Explanatory Gap” (vol. 06, no. 2) Sara Worley refers closely to Dihlhay’s theory. According to Worley, Harman tries to reduce the problem of the explanatory gap to the issue of various explanatory frameworks. She asks, “Can we learn what it’s like to be a bat by living with bats?” Worley claims that “the peculiarly first person phenomenal experience” cannot be acquired by the sort of learning needed in acquiring a new theoretical framework. But isn’t the whole point of Harman’s functionalism of concepts [Harman 1993, see bibliography in Worley’s paper] to show just this, and, thereby, that accepting first person experience is the best explanation of this very difficulty, encountered by reductive functionalism(s)?

The main bulk of discussion is devoted to L. Floridi’s recent article, “Understanding Information Ethics” (vol. 07, no. 1). We are happy to present commentaries by Antonino Vaccaro on further perspectives of information ethics as macroethics, and John P. Sullins on the status of virtual worlds. Elizabeth A. Buchanan focuses on some social and ethical implications of Floridi’s approach to the infosphere while Samir Chopra ponders on what a truly “frictionless infosphere” would involve. Commentaries by E.H. Spence and Terry Bynum are slated for the next issue.

Following Michael Byron’s objective (“Notes from the Incoming Chair,” vol. 07, no. 1) to keep a balance between the two constituencies of this committee—those working on issues such as AI, philosophical impact of computer science on the issues of ontology, epistemology, philosophy of mind, and related domains; and those primarily interested in the use of computers in education—we devote the next section to “Discussion Papers on Online Education and Philosophy.” The section begins with a paper by Marvin Croy, deriving from his recent APA conference presentation devoted to evaluation and perspectives of teaching machines. Croy’s presentation made quite a splash on the net; it will be interesting to have a first-hand account of his main ideas. We look forward to having Harriet Baber’s paper, based on contribution to the same session, in the next issue of this Newsletter.

A nice follow-up on the paper by Croy, as well as the one by Halpin, is Marie Lyle’s “When They Program Us.” It is a refreshing reflection on online education in the broader context of philosophy of mind, relying on the work of A. Clarke. Its earlier version was presented at the ’06 NA-CAP. In his paper “The Blended Classroom,” Pat Manfredi, one of the veterans of online education in philosophy, provides tips on how to use the Internet, with moderation, in on-campus and blended courses. Manfredi’s paper will be particularly helpful to teachers in colleges that do not support online learning platforms; it is based on a 2006 presentation at the session organized by this committee with the APA Central Division. We are glad to also have a paper by Derek Matravers presenting the philosophy program at the British Open University, which provides education within the traditional and proven model of distance education (with local tutors and all) enhanced by the use of the Internet. While in general those courses are not available in the United States, it is refreshing to know that, in Britain, Open University offers undergraduate and graduate degrees in philosophy while the department employs a number of high-caliber philosophers teaching exclusively distance education courses.
The Committee will also host a special session on Pedagogical Developments in Philosophy and Computers, chaired by Peter Boltuc, which includes papers by Patrick Suppes, Marvin Croy, and Peter Boltuc. The session is a follow up on a successful session devoted to the same issues in 2006 that included the same speakers.

I would like to thank the Committee members for helping make my transition as incoming chair of the Committee smooth. I look forward to continuing our work together.

**ARTICLES**

*The Shrinking Difference Between Artifacts and Natural Objects*

Lynne Rudder Baker  
*University of Massachusetts–Amherst*

Artifacts are objects intentionally made to serve a given purpose; natural objects come into being without human intervention. I shall argue that this difference does not signal any ontological deficiency in artifacts qua artifacts. After sketching my view of artifacts as ordinary objects, I'll argue that ways of demarcating genuine substances do not draw a line with artifacts on one side and natural objects on the other. Finally, I'll suggest that philosophers have downgraded artifacts because they think of metaphysics as resting on a distinction between what is “mind-independent” and what is “mind-dependent.” I'll challenge the use of any such distinction as a foundation for metaphysics.

**Artifacts as Ordinary Objects**

Artifacts should fit into any account of ordinary objects for the simple reason that so many ordinary objects are artifacts. We sleep in *beds*; we eat with *knives and forks*; we drive *cars*; we write with *computers* (or with *pencils*); we manufacture *nails*. Without artifacts, there would be no recognizable human life.

On my view—I call it “the Constitution View”—all concrete objects, except for “simples” if there are any, are ultimately constituted by sums (or aggregates) of objects. Technical artifacts—artifacts made to serve some practical purpose—are, like nonartifacts, constituted by lower-level entities. Constitution is a relation of unity-without-identity. Unlike identity, constitution is a contingent and time-bound relation. To take a simple-minded example, consider a wooden rod and a piece of metal with a hole just slightly bigger than the diameter of the rod. When the aggregate of the rod and the piece of metal are in certain circumstances (e.g., when someone wants to make a hammer and inserts the rod into the hole in the metal), a new object—a hammer—comes into being. Since the rod and the piece of metal existed before the hammer did, the relation between the aggregate of the rod and the piece of metal and the hammer is not identity. It is constitution.

Typically, artifacts are constituted by aggregates of things. But not always: a paperclip is constituted by a small piece of thin wire; and a $50 bill is constituted by a piece of paper. Nevertheless, the piece of thin wire and the piece of paper themselves are constituted by aggregates of molecules, which in turn are constituted by aggregates of atoms. So, even those artifacts (like paperclips) that are constituted by a single object are, at a lower level, constituted by aggregates of atoms. For simplicity, I'll consider artifacts to be constituted by aggregates of things, not by a single object. Any items whatever are an aggregate. The identity conditions of aggregates are simple:
aggregate x is identical to aggregate y just in case exactly the same items are in aggregate x and aggregate y.

**Differences Between Artifacts and Natural Objects**

Technical artifacts have proper functions that they are designed and produced to perform (whether they successfully perform their proper functions or not). Indeed, the general term for a kind of artifact—e.g., polisher, scraper, life preserver—often just names the proper function of the artifact. An artifact has its proper function essentially: the nature of an artifact lies in its proper function—what it was designed to do, the purpose for which it was produced. Moreover, artifacts have their persistence conditions in virtue of being the kind of artifact that they are. Put an automobile in a crusher and it—it—goes out of existence altogether. The metal and plastic cube that comes out of the crusher is not the same object (your old clunker of a car) that went in. Since artifacts have intended functions essentially, they are what I call “intention-dependent” or “ID” objects: they could not exist in a world without beings with propositional attitudes.

Natural objects differ from artifacts in at least three ways:

1. Artifacts (and not natural objects) depend ontologically—not just causally—for their existence on human purposes. (2) Relatedly, artifacts are “intention-dependent” (ID) objects that could not exist in a world without minds. Natural objects, which can be deployed to serve human purposes, would exist regardless of human intentions or practices. (3) Artifacts (and not natural objects) essentially have intended proper functions, bestowed on them by beings with beliefs, desires, and intentions.

**The Ontological Status of Artifacts**

Many important philosophers—from Aristotle on—hold artifacts ontologically in low regard. Some philosophers have gone so far as to argue that “artifacts such as ships, houses, hammers, and so forth, do not really exist.” Artifacts are thought to be lacking in some ontological way: they are considered not to be genuine substances. Although the notion of substance is a vexed one in philosophy, what I mean by saying that things of some kind—e.g., hammers, dogs, persons—Fs in general—are genuine substances is that any complete account of what there is will have to include reference to Fs. I shall argue that there is no reasonable basis for distinguishing between artifacts and natural objects in a way that renders natural objects as genuine substances and artifacts as ontologically deficient.

I shall consider five possible ways of distinguishing between natural objects and artifacts, all of which are mentioned or alluded to by David Wiggins. On none of these, I shall argue, do natural objects, but not artifacts, turn out to be genuine substances. Let the alphabetic letter “F” be a placeholder for a name of a type of entity.

1. Fs are genuine substances only if Fs have an internal principle of activity.
2. Fs are genuine substances only if there are laws that apply to Fs as such, or there could be a science of Fs.
3. Fs are genuine substances only if whether something is an F is not determined merely by an entity’s satisfying a description.
4. Fs are genuine substances only if Fs have an underlying intrinsic essence.
5. Fs are genuine substances only if the identity and persistence of Fs is independent of any intentional activity.

Let us consider (1) through (5) one at a time.

(1) The first condition—Fs are genuine substances only if Fs have an internal principle of activity—has its source in Aristotle. Aristotle took this condition to distinguish objects that come from nature (e.g., animals and plants) from objects that come from other efficient causes (e.g., beds). But this condition does not rule in natural objects and rule out artifacts as genuine substances. A piece of gold is a natural object, but today, we would not consider a piece of gold (or any other chemical element) to have an internal principle of change; conversely, a heat-seeking missile is an artifact, but it does have an internal principle of activity. So, the first condition does not distinguish artifacts from natural objects.

(2) The second condition—Fs are genuine substances only if there are laws that apply to Fs as such, or there could be a science of Fs—also allows artifacts to be genuine substances. Engineering fields blur the line between natural objects and artifacts. Engineering schools have courses in materials science (including advanced topics in concrete), traffic engineering, transportation science, computer science—all of which quantify over artifacts. Since something’s being of an artifactual kind (e.g., computer) does not preclude a science of it, the second condition does not make artifacts less than genuine substances.

(3) The third condition is semantic: Fs are genuine substances only if whether something is an F is not determined merely by an entity’s satisfying a description. Demonstrative reference is supposed to be essential to natural-kind terms. The reference of natural-kind terms is said to be determined indexically; the reference of artifactual-kind terms is said to be determined by satisfying a description.

Membership in a natural kind, it is thought, is not determined by satisfying a description, but rather by relevant similarity to stereotypes. The idea is this: First, Fs are picked out by their superficial properties (e.g., quantities of water are clear liquids, good to drink, etc.). Then, anything that has the same essential properties that the stereotypes have is an F. So, natural kinds have “extension-involving sortal identifications.” By contrast, artifactual terms (like those I used earlier—“beds,” “knives and forks,” “cars,” “computers,” “pencils,” “nails”) are said to refer by satisfying descriptions: “A clock is any time-keeping device, a pen is any rigid ink-applying writing implement and so on.”

I do not think that this distinction between how words refer captures the difference between natural objects and artifacts. The distinction between referring indexically and referring by description, with respect to natural kind terms, is only a matter of the state of our knowledge and of our perceptual systems. However gold was originally picked out (e.g., as “stuff like this”), now we can pick it out by [what are taken to be] its essential properties: for example, Gold is the element with atomic number 79. Not only might natural kinds satisfy descriptions, but also we may refer to artifacts in the absence of any identifying description. For example, archeologists may believe that two entities are both artifacts of the same kind, without having any identifying description of the kind in question. (Were they used in battle or in religious rituals?)

Thus, the third condition—Fs are genuine substances only if whether something is an F is not determined merely by an entity’s satisfying a description—does not distinguish natural kinds from artifactual kinds, nor does it rule out artifacts as genuine objects.

(4) The fourth condition—Fs are genuine substances only if Fs have an underlying intrinsic essence—also fails to distinguish natural from artifactual kinds. Although some familiar natural kinds—like water or gold—have underlying intrinsic essences, not all do. For example, wings (of birds and insects), mountains, and planets are all natural kinds, but none of them has an
underlying intrinsic essence. Their membership in their kinds is not a matter of underlying intrinsic properties. Something is a wing, mountain, or planet not in virtue of what it is made of, but in virtue of its relational properties. For that matter, something is a bird or an insect in virtue of its relational properties—its genealogical lineage. So, having an underlying intrinsic essence does not distinguish natural objects from artifacts.

(5) The fifth condition—Fs are genuine substances only if the character of F is independent of any intentional activity—is the most interesting. According to some philosophers, the “character of [a] substance-kind cannot logically depend upon the beliefs or decisions of any psychological subject.”15 Unlike the first four conditions, the fifth does distinguish between artifactual and natural kinds. An artifact’s being the kind of thing that it is depends on human intentions. Conceding that the necessity of intention is a difference between an artifact and a natural object, I ask: Why should this difference render artifacts deficient?

If you endorse what Jaegwon Kim has called “Alexander’s Dictum”—To be real is to have effects—there is no doubt that artifacts are real. When automobiles were invented, a new kind of thing came into existence: and it changed the world. Considering the world-changing effects of the automobile (and countless other kinds of artifacts), artifacts have as strong a claim to ontological status as natural objects.

What generally underlies the fifth condition, I believe, is an assumption that Fs are genuine substances only if conditions of membership in the substance-kind are set “by nature, and not by us.”16 But it is tendentious to claim that the existence of artifacts depends not on nature, but on us.17 Of course the existence of artifacts depends on us: but we are part of nature. It would be true to say that the existence of artifacts depends not on nature—as-if-we-did-not-exist, but depends on nature-with-us-in-it. Since nature has us in it, this distinction (between nature-as-if-we-did-not-exist and nature-with-us-in-it) is no satisfactory basis for a verdict of ontological inferiority of artifacts.

The Insignificance of the Mind-Independence/Mind-Dependence Distinction

There is a venerable—but, I think, theoretically misguided—distinction in philosophy between what is mind-independent and what is mind-dependent. Anything that depends on our conventions, practices, or language is mind-dependent (and consequently downgraded by those who rest metaphysics on a mind-independence/mind-dependence distinction). All ID objects, including all artifacts, are by definition mind-dependent, inasmuch as they could not exist in a world without beings with beliefs, desires, and intentions. Nothing would be a carburetor in a world without intentional activity.18 The mind-independent/mind-dependent distinction is theoretically misguided because it is used to draw an ontological line in an unilluminating place. It puts mind-independent insects and galaxies on one side, and mind-dependent afterimages and artifacts on the other.

A second reason that the mind-independent/mind-dependent distinction is unhelpful is that advances in technology have blurred the difference between natural objects and artifacts. For example, so-called “digital organisms” are computer programs that (like biological organisms) can mutate, reproduce, and compete with one another.19 Or consider “robo-rats”—rats with implanted electrodes that direct the rats’ movements.20 Or, for another example, consider what one researcher calls “a bacterial battery”:21 these are biofuel cells that use microbes to convert organic matter into electricity. Bacterial batteries are the result of a recent discovery of a micro-organism that feeds on sugar and converts it to a stream of electricity. This leads to a stable source of low power that can be used to run sensors of household devices. Finally, scientists are genetically engineering viruses that selectively infect and kill cancer cells and leave healthy cells alone. Scientific American referred to these viruses as “search-and-destroy missiles.”22 Are these objects—the digital organisms, robo-rats, bacterial batteries, genetically engineered viral search-and-destroy missiles—artifacts or natural objects? Does it matter? I suspect that the distinction between artifacts and natural objects will become increasingly fuzzy; and, as it does, the worries about the mind-independent/mind-dependent distinction will fade away. More particularly, as the distinction between natural objects and artifacts pales, the question of the ontological status of web-based objects, for example, becomes more acute.

Conclusion

No one who takes artifacts of any sort seriously, ontologically speaking, should suppose that metaphysics can be based on a distinction between mind-independence and mind-dependence. In any case, technology will continue to shrink the distinction, and, with it, the distinction between artifacts and natural objects.23

Endnotes

1. An earlier version of this paper was presented to the Society of Philosophy and Technology, Chicago APA, April 20, 2007.

1. There is a lot of literature on functions. For example, see Crawford L. Elder, “A Different Kind of Natural Kind,” Australasian Journal of Philosophy 73 (1995): 516-31. See also Pieter E. Vermaas and Wybo Houkes, “Ascribing Functions to Technical Artifacts: A Challenge to Etiological Accounts of Functions,” British Journal for the Philosophy of Science 54 (2003): 261-89. As Vermaas and Houkes point out, some philosophers take the notion of biological function to be basic and then try to apply or transform theories of biological function (which since Darwin are non-intentionalist, reproduction theories) to artifacts. I believe that Vermaas and Houkes are entirely correct to liberate the theory of artifacts from the notion of function in biology.


3. More precisely, a nonderivative artifact has its proper function essentially. The constituter of an artifact inherits the nonderivative artifact’s proper function and thus has it contingently (as long as it constitutes the nonderivative artifact).


5. All the conditions either follow from, or are part of, the basic distinction that Wiggins draws between natural objects and artifacts. There is a complex condition that natural objects allegedly satisfy and artifacts do not: “...a particular constituent x belongs to a natural kind, or is a natural thing, if and only if x has a principle of activity founded in lawlike dispositions and propensities that form the basis for extension-involving sortal identification (s) which will answer truly the question ‘what is x?’” According to Wiggins, natural objects satisfy this condition and artifacts do not. David Wiggins, Sameness and Substance Renewed (Cambridge: Cambridge University Press, 2001), 89. I am not claiming that Wiggins denies that there exist artifacts, only that he distinguishes between natural and artifactual kinds in ways that may be taken to imply the ontological inferiority of artifacts.

6. A substance has “within itself a principle of motion and stationariness (in respect of place, or of growth and decrease, or by way of alteration).” Aristotle, Physics 192b8-23.

7. This claim is similar to the notion that natural-kind terms, but not artificial-kind terms, are rigid designators. (A rigid designator has the same referent in every possible world.) However, what makes the difference between “whale” and...
"bachelor" is not that only the former is rigid. Rather, only the former term "has its reference determined by causal contact with paradigm samples of the relevant kind." There is no reason that the terms cannot both be rigid. See Joseph LaPorte, "Rigidity and Kind," Philosophical Studies 97 (2000): 304.

Although Wiggins is an Aristotelian, this is not Aristotle's view. For Aristotle, nominal definitions are reference fixers, used to identify objects for scientific study; they contain information that a scientist has before having an account of the essence of the objects. Real definitions are discovered by scientific inquiry and give knowledge of the essences of objects identified by nominal definitions. Nominal and real definitions are not accounts of different types of entities. Rather, they are different types of accounts of the same entities. Members of a particular natural kind have the same essence (underlying structure). See Robert Bolton, "Essentialism and Semantic Theory in Aristotle: Posterior Analytics, II, 7-10," The Philosophical Review 85 (1976): 514-44.

E.g., Wiggins, Sameness and Substance Renewed, pp. 11-12.

Ibid., 89.

Ibid., 87.

Aristotle would agree with me on this point, I believe. His reason for downgrading artifacts ontologically is that artifacts have no natures in themselves.

Moreover, indexicality should not be confused with rigidity, which does not concern how a term gets connected to a referent. For criticism of Putnam's confusion of the causal theory of reference and indexicality, see Tyler Burge, "Other Bodies," in Thought and Object, edited by Andrew Woodfield (Oxford: Oxford University Press, 1982), 97-120.

Joseph LaPorte also holds that some kind expressions (both natural and artificial) designate rigidly, and some designate nonrigidly. See his "Rigidity and Kind," Philosophical Studies 97 (2000): 293-516.


In Chapter One of The Metaphysics of Everyday Life (Cambridge: Cambridge University Press, 2007), I argued that a distinction between what depends on nature and what depends on us is neither exclusive nor exhaustive.


The New York Times, September 18, 2003. The lead researcher, Derek Lovley, who coined the term "bacterial battery," is a microbiologist at the University of Massachusetts-Amherst.

Email update from Scientific American, September 23, 2003.


---

**Philosophical Engineering: Towards a Philosophy of the Web**

**Harry Halpin**

University of Edinburgh

**Abstract**

The Web is commonly considered the most significant computational phenomenon to date. However, the Web itself has received scant attention from philosophy, being best regarded as a mere engineering artifact. Furthermore, the efforts to evolve the Web into the Semantic Web are viewed with suspicion by most philosophers as a return to Cartesian artificial intelligence. I argue that these widely held viewpoints are incorrect, and that the Web succeeds because of its design principles that distinguish it from both previous hypertext systems and knowledge representation systems in classical artificial intelligence. Furthermore, the Web embodies the logical conclusion of Clark's Extended Mind thesis since it allows multiple individuals to access and manipulate the same representation, so offering the ultimate in cognitive scaffolding. This undermines the notion of individual intelligence at the heart of Cartesian artificial intelligence and presents a challenge to the role of representations as given in the recent wave of neo-Heideggerian focus on embodiment. Taking the Web seriously moves the primary focus of philosophy away from the role, or lack thereof, of internal representations to external representations. The Web is then properly understood as the creation and evolution of external representations in a universal information space. Berners-Lee calls this "philosophical engineering," and it has surprising connections to neo-Fregeanism, anti-realism, and other long-standing philosophical debates.

**The Web as Philosophy**

The reigning model of computation usually considered by philosophers debating artificial intelligence is the lone Turing Machine manipulating representations in an existential void. The most controversial proposition of computer science that captures the imagination of philosophy is whether or not this model can be authentically intelligent. Under heavy attack from philosophers ranging from Dreyfus to Wheeler, this hypothesis of classical artificial intelligence is showing wear and tear, if not total defeat. From the ashes of what has been termed "Good Old Fashioned Artificial Intelligence" has arisen a neo-Heideggerian theory of embodiment as based on work in dynamical systems, robotics, and artificial life. This line of work is inspired by the claim of Rodney Brooks that intelligence does not require representations (1991). While impressive, the vast majority of work in computing is not the building of embodied robots capable of simulating emotions. While the sheer ambition of embodied artificial intelligence grabs the imagination of philosophers, a revolution of everyday life has emerged due to the advent of the Web. The Web, and the complex web of human and computer interaction it engenders, is already supplanting the isolated digital computer in everyday life, and so will eventually supplant it as the basis for philosophical analysis of the nature of computation and the mind. We argue that the Web represents a fundamental turning point not just in computational systems but in philosophy as well, for the Web signals a return of representationalism of a distinctly different kind than that formerly theorized about by artificial intelligence, "embodied" or otherwise.

In his "One Billion Lines of C++," Brian Cantwell Smith notes that the models of computing used in philosophical debates over intelligence, representation, embodiment, and...
consciousness ignore the majority of existing computers by framing the debate as if it were between logic-based symbolic reasoners and some alternative ranging from neural networks to epigenetic robotics (1997). As Smith points out, “It is impossible to make an exact estimate, but there are probably something on the order of 10, or one hundred billion lines of C++ in the world. And we are barely started. In sum: symbolic AI systems constitute approximately 0.01% of written software” (Smith, 1997). The same small fraction likely holds true of “non-symbolic AI” computational systems such as robots, artificial life, and connectionist networks. Numbers by themselves hold little intellectual weight, for one could always argue that the vast majority of computational systems are simply philosophically uninteresting. However, trends in philosophy contradict this intuition. The turn in philosophy away from artificial intelligence and linguistics to what has been termed a “philosophy of information” demonstrates that this wider class of computational systems is finally having an impact (Floridi 2004).

The most significant computational system to date is the World Wide Web, described by its inventor Tim Berners-Lee as “a universal information space” (Berners-Lee 1998). Due to its hegemonic role today, understanding the principles that distinguish the Web from other computational systems should be a goal of the philosophy of information. To articulate what I term the “philosophy of the Web,” one needs to first clarify whether or not the Web is continuous or in conflict with the development of current trends in philosophy, cognitive science, and artificial intelligence. This is not an easy endeavor at first glance, for the Web in nothing if not a robustly representational system, and a large amount of research on the Web focuses on how to enable increasingly powerful and flexible forms of representations. One development Berners-Lee calls the “Semantic Web” consists of standardizing the logical encoding of representations so that they may be directly usable by machines. There is suspicion among many of the now successful rebels against classical artificial intelligence that “Good Old Fashioned” artificial intelligence has not disappeared but has been instead born anew on the Semantic Web.

Note that we do not present a fully developed philosophical theory of the Web, which would be far out of scope, but a broad picture of the philosophy of the Web, an intellectual endeavor that needs to be fleshed out and thoroughly argued in other work. Yet, in order to have detailed arguments over a philosophy of the Web, a precis that relates a new developing picture to more widely known and accepted paradigms is necessary.

Beyond Neo-Heideggerian Embodiment

It is difficult to summarize the hypothesis made by proponents of embodied artificial intelligence, yet this must be done in order to understand if this hypothesis is in fact continuous with the Web. In contrast, in defining classical artificial intelligence, Smith stated the Knowledge Representation Hypothesis:

Any mechanically embodied intelligent process will be comprised of structural ingredients that a) we as external observers naturally take to represent a propositional account of the knowledge that the overall process exhibits, and b) independent of such external semantical attribution, play a formal but causal and essential role in engendering the behavior that manifests that knowledge. (Smith 1985)

One compelling story put forward by the philosophers of embodied artificial intelligence is by Michael Wheeler (2005). Instead of phrasing a negative critique of classical artificial intelligence like Hubert Dreyfus, Wheeler turns his Heideggerian analysis into a positive program for embodied artificial intelligence. In order to contrast classical artificial intelligence with its embodied alternative, Wheeler produces three Cartesian claims that he believes underlie classical artificial intelligence (2005):

1. The subject-object dichotomy is a primary characteristic of the cognizer’s ordinary epistemic situation.
2. Mind, cognition, and intelligence are to be explained in terms of representational states and the ways in which such states are manipulated and transformed.
3. The bulk of intelligent human action is the outcome of general purpose reasoning processes that work by retrieving just those mental representations that are relevant to the present behavioral context and manipulating and transforming those representations in appropriate ways as to determine what to do.

Wheeler then states that “word on the cognitive-scientific street is that classical systems have, by and large, failed to capture in anything like a compelling way, specific styles of thinking at which most humans naturally excel” (2005). Wheeler then contrasts the Cartesian assumptions of classical artificial intelligence with Heidegger in order to formulate the philosophical principles of embodied artificial intelligence (Wheeler 2005):

1. **The primacy of online intelligence**: The primary expression of biological intelligence, even in humans, consists not in doing math or logic, but in the capacity to exhibit...online intelligence...a suite of fluid and flexible real-time adaptive responses to incoming sensory stimuli.
2. **Online intelligence is generated through complex causal interactions in an extended brain-body-environment system**: Online intelligent action is grounded not in the activity of neural states and processes alone, but rather in the complex causal interactions involving not only neural factors, and also additional factors located in the non-neural body and the environment.
3. **An increased level of biological sensitivity**: Humans and animals are biological systems—and that matters for cognitive science.
4. **A dynamical systems perspective**: Cognitive processing is fundamentally a matter of state space evolution in certain kinds of dynamical systems.

These assertions we call the neo-Heideggerian framework of embodied cognitive science. Wheeler and other philosophers then return to the old question of whether or not there is room for internal representations in cognitive science. Wheeler argues that there is a limited role for internal representations to play, and work in embodied cognitive science should no longer only tout that the world is its own best model (Brooks 1991). Even work in robotics has shown that internal representations are incredibly important when “the world is not enough” and that training via simulation can be effective replacement for an often dangerous world (Grush 2003). The world may be its own best model, but it can get you killed. Therefore, some limited form of representations may have an evolutionary advantage for those that can use them (Grush 2003). This is a small victory at best for representations, for representations are cast from their long-standing spotlight in classical artificial intelligence into a secondary role in embodied cognitive science. Can representations in the usual sense fit within this neo-Heideggerian framework?

In the details, Wheeler is not as far from us as he may appear, even though we have intellectual projects that are
quite far apart. One of the most thoughtful neo-Heideggerian philosophers, Wheeler holds that decoupling could be sufficient for representation, although not necessary since there are cases of on-line intelligence where decoupleability does not hold but can nonetheless be explained by representations and so Wheeler is far less skeptical about representations than others like Dreyfus. While Wheeler argues for homuncularity and arbitrariness, his task is to explain the use of internal representations in explanations (2007). For the time being, one can be agnostic about the role of arbitrariness and homuncularity in explanation. What we need is a notion of what a representation is, a definition that applies to both “internal” and “external” representations, not conditions for a representational explanation in cognitive science.

What are Representations?

The very idea of representation is usually left under-defined as a vague “standing-in” intuition. The classic definition of a symbol from the Physical Symbol Systems Hypothesis is the genesis of this intuition regarding representations (Newell 1980):

An entity X designates an entity Y relative to a process P, if, when P takes X as input, its behavior depends on Y.

There are two keys to this definition. First, the concept of a representation is grounded in the behavior of a process. Thus, what precisely counts as a representation is never context-free. Second, the representation simulates action at a distance on what is being represented, “This is the symbolic aspect, that having X (the symbol) is tantamount to having Y (the thing designated) for the purposes of process P” (Newell 1980). This definition seems to have an obvious point of conflict with the neo-Heideggerian agenda, for it reflects a “subject-object dichotomy” due to its presupposition of three ontologically distinct entities. To return to the definition put forward by the Physical Symbol Systems Hypothesis, let us call the process P the “subject” that is using the representation. Let us call X the “representation” and Y the “object.” Therefore, the subject-object dichotomy is actually present three times in the Physical Symbol Systems Hypothesis: the subject, the object, and the representation. The dichotomy is present between the subject and its object, the object and its representation, and even perhaps the subject and its representation. To the extent that these distinctions are ontologically held a priori, then the definition is hopelessly Cartesian.

The only way to escape this Cartesian trap is to give a description of how representations arise without the a priori subject-object dichotomy. This is precisely what Brian Cantwell Smith proposes through his process of registration (1996). Smith starts with the classic example of a frog tracking a gadfly across the sky. The frog sees the gadfly and begins tracking the gadfly with its eyes as it flies. The frog and the gadfly are both physically connected via light-rays. Borrowing an analogy from physics, Smith notes that everything is composed of non-distinct fields of energy, so it would be ontologically wrong to talk about a frog, a gadfly, and light as individual objects. All that exists are fields that are physically connecting and disconnecting from each other to greater or lesser degrees. At the moment of tracking, one can speak of the frog and gadfly as a single connected field. When the gadfly goes behind a tree, and then when the fly emerges from the other side of the tree, the eyes of the frog are not focused on the point the gadfly was at before it went behind the tree, but the point the gadfly would be at if it continued on the same path. Although this is a simple trick capable of being accomplished without any full-blooded internal representation, Smith believes this sort of behavior builds the foundation of intentionality and so for the eventual emergence of representation. In the language of fields, the fields separate into an o-region disconnected from the s-region. The s-region is distinguished from the o-region by virtue of the s-region’s attempt to “track” the o-region, its attempt to remain connected in “a long-distance coupling against all the laws of physics” (Smith 1996). The s-region eventually stabilizes as the subject and the o-region as an object, with considerable work on at least the subject’s side. This work manifests itself as the creation of a representation that the subject maintains of the object. The subject, the representation of the object, and the object itself are the final result of registration. Furthermore, the subject and object are not a priori distinct, but co-constitute each other. The distinction between subject and object is given by the use of a representation of the object by the subject.

In order to explicate what precisely the subject must possess in order to track the object via a representation, we rely on Rocha and Nordjik’s notion of dynamically incoherent memory (2005). “Dynamically incoherent” means the memory is not changed by any dynamic process it initiates or encounters. The term “dynamically incoherent” is misleading, for most people would say “dynamically incoherent” actually means the maintaining of coherence against “the vagaries and vicissitudes, the noise and drift, of earthly existence,” as Haugeland would say (1981). In other words, somehow the object must have a memory that allows it to store a representation for some period of time with a degree of fidelity and reliably. Of course, this is precisely what digital memory is good for.

However fuzzy the details of Smith’s story about representations may be, what is clear is that instead of positing the subject, object, and representation a priori, they are introduced as products of a temporal process. This process is at least theoretically “non-spooky” since the entire process is capable of being grounded out in physics without any “spooky” action at a distance. To be grounded out in physics, all changes must be given in terms of contact in space-time, or, in other words, “local” contact. In this way, representations are “a way of exploiting local freedom or slop in order to establish coordination with what is beyond effective reach” (Smith 1996). In order to clarify Smith’s registration and improve the definition of the Physical Symbol Systems Hypothesis, we consider this entire story to be a representational cycle (Halpin 2006):

1. **Presentation**: Process S is in local contact with process O. S is the s-region that evolves into the subject, while O is the o-region that evolves into the object.
2. **Input**: The process S is in local contact with coherent memory R. An input procedure of S puts R in correspondence with some portion of process O. This is entirely non-spooky since S and O are in local contact with R. R evolves into the representation.
3. **Separation**: Processes O and S change in such a way that the processes are non-local.
4. **Output**: Due to some local change in process S, S uses its local contact with R to initiate local behavior that depends on R for success.

So we have constructed an ability to talk about representations while not presupposing that intelligent behavior depends on internal representations or that representations exist a priori at all. Representations are only needed when the relevant intelligent behavior requires some sort of distal co-ordination. In this manner, representations, if not representationalism, are continuous with the neo-Heideggerian agenda. Representations are not a Cartesian metaphysical assumption, but arise over time in a way that is not only coherent with physics but also with the neo-Heideggerian program. Representations exist as part of a rich temporal dynamic that does not presuppose a Cartesian
subject-object dichotomy, instead being based on contingent and temporary object-subject dichotomies.

From the Extended Mind to External Representationalism

The success of Smith’s argument lies also in its ability to phrase the creation of representations that are neither internal or external to a particular subject. Furthermore, just because the coherent memory that serves as a representation is at some point in local contact with a subject, it does not mean that it must always be tied to the subject. In other words, the representation does not have to be “in the head” of the subject. In this case, we can divide the world of representations into two types, those that are internal to the subject and those that are external to the subject. Although this argument undermines the Cartesian use of “internal” and “external” in general, for the rest of this argument we will just use the term “internal representation” to designate representations that are implemented biologically inside a human subject as traditionally defined by the bounds of the skin. We will use “external representation” to designate representations that are implemented outside what is traditionally considered outside the biological body of a human subject.

One of the tenets of the neo-Heideggerian program put forward by Wheeler is that “online intelligence is generated through complex causal interaction in an extended brain-body-environment system” (2005). There is no reason why representations cannot be part of that environment, since in lieu of Smith’s story of registration we have rephrased representations as not necessarily being internal. In fact, we can remain agnostic as regards to the possibility of whether or not internal representations are necessary or even used by a human subject, leaving the debate over internal representations to a purely empirical question best left to cognitive science. These representations in the environment can even be crucial to intelligence, leading to what Clark and Chalmers call “an active externalism, based on the active role of the environment in driving cognitive processes” in their Extended Mind thesis (1999).

To explain their Extended Mind thesis, Clark and Chalmers introduce us to the charming Otto, a man with an impaired memory who navigates about his life via the use of his notebook (1999). Otto is trying to navigate to the Metropolitan Museum of Modern Art in New York City from his house, and uses his notebook as a surrogate memory in order to discover the address. Let’s clarify by adding to the original example that in order to successfully arrive at the museum, Otto needs a map which is in some correspondence with the world he must navigate through in order to successfully get to the museum. We can imagine Otto has a map in his notebook to the Museum of Modern Art made for the precise purpose of navigating individuals to the museum. Otto can get to the museum with the map, and without it he would be lost. It is hard to deny that a map is representational in the sense we have presented above. In this regard, external representations do exist in the environment of an agent and can drive the cognitive processes of an agent in a similar fashion to the way that classical artificial intelligence assumed internal representations did. Interestingly enough, Clark and Chalmers point out that if external factors are driving the process, then they deserve some of the credit: “If, as we confront some task, a part of the world functions as a process which, were it done in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process” (1999). In this regard, the Extended Mind thesis undermines the strict division between internal and external of the subject and its representation in a way that is compatible with the neo-Heideggerian philosophical framework.

The Extended Mind thesis in combination with the representational cycle explains the widespread success of digital computing. The analog memories of humans are notoriously bad at maintaining any sort of coherence in front of the push and pull of everyday life. This means that humans who discover a way to maintain representations outside of their biological skin may prove to have an evolutionary advantage over those who don’t have such an ability. The maintenance of coherent memory requires considerable work. In digital memory, this is exemplified in how voltages must be carefully regulated to maintain the coherence of computer memory. With the advent of digital memory, where representations can be safely encoded and recoded at a level of abstraction that is resistant to change, a new medium for representations has been created that is in general more suited for representations than biological memory. This may explain the strange fascination of classical artificial intelligence for digital computing and internal representationalism, even though many representation-heavy tasks studied by classical artificial intelligence constitute “psychological arenas in which most humans perform rather badly, and in which most other animals don’t perform at all” (Wheeler 2005). We might want to amend this observation. Classical artificial intelligence specialized in tasks that humans perform rather badly at when not aided by machines with digital memory. With the advent of digital memory and computation humans can now be successful at all sorts of things that humans and other animals without the cognitive scaffolding of digital technology would fail at normally.

To press upon the Extended Mind thesis, imagine the world to be inhabited by multiple subjects that can access the same representation. In almost all the original examples that Clark and Chalmers use in the Extended Mind thesis, they use a lone person sitting in front of a computer screen. This ignores the use of multiple people using the computer as a collaborative and communications tool. The obvious example would be two people using the Internet to both share a single representation. One could imagine Otto trying to find his way to the Museum of Modern Art, and instead of a notebook having a mobile telephone with access to a web page that has a map. One could also imagine Inga, who may not have Alzheimer’s but nonetheless cannot remember her way to the Museum, having access to the same map via her personal digital assistant’s Internet access. Since they are sharing the representation and their behavior is normatively successful based on its use, Inga and Otto can be said to partially share the same cognitive state.

To push the Extended Mind thesis even further, imagine not only that Otto and Inga are using a web page with a map, but a web page that allows users to add annotations to the map. The web page is updated with the annotations for all to use in near real time. Inga, noticing that the main entrance to the Museum of Modern Art is closed temporarily due to construction and so the entrance has moved over a block, adds this annotation to the map, correcting an error as regards where entrance should be. Luckily, with this change to the map Otto can now find the entrance to the museum, while without it he would have been hopelessly lost. This active manipulation of a representation lets Inga and Otto partially share a dynamic cognitive state and collaborate for their greater collective success. Their shared cognitive process is functioning not via telepathy but via shared external representations that are universally accessible over the Web. Clark and Chalmers agree that cognition can be socially extended: “What about socially extended cognition? Could my mental states be partly constituted by the states of other thinkers? We see no reason why not, in principle” (1999). This socially extended cognition can be accomplished via shared external representations. In fact, it is precisely to these ends that
the constant work on the Web is heading, especially the more interactive and collaborative Web technologies.

**The Principles of the Web**

Unlike classical artificial intelligence’s emphasis on internal representations, the Web is primarily concerned with external representations. In contrast to the champions of the embodied mind, the Web is concerned with exactly what digital, external representations can be used to maximize intelligence. It is perhaps an irony that the embodied mind is making such gains while human society and intelligence seem increasingly disembodied. The value of external representations comes with their accessibility, for an external representation that is unaccessible cannot enable online intelligence. It is precisely in order to solve this problem—the problem of the accessibility of external representations—that Tim Berners-Lee proposed the Web as “a universal information space” (URI) (1998).

The primary advantage of the Web is that representations can be accessed via a URI so that it may link with any other representation. Furthermore, a URI identifies a “resource,” which may in turn host multiple representations of the same object. From this foundation in URIs, a number of standardized protocols for accessing representations and formats for encoding representations provide the foundation of the Web. From the perspective of philosophy, it is interesting that traditional philosophy has in general taken inspiration from artificial intelligence, although it is only one branch of research in computing. Not as well known as the lineage of artificial intelligence, the Web has an historical lineage in Douglas Engelbart’s “Human Augmentation Framework” (1962), a parallel and sometimes competing program to artificial intelligence. The Web inherits much inspiration and even design from the work of Engelbart’s NLS (onLine System). More and more recently his system and ideas have begun to receive attention from philosophers (Cooksey 2006).

What would be philosophically important would be if certain kinds of external representations were found to be amendable to facilitating intelligence. To answer this question, the exact characteristics of the external representations used by the Web need to be articulated. While their articulation requires far more detailed explanation, the principles of the Web may be defined tentatively as:

1. **The Principle of Universality**: Everything can be given a URI, and therefore everything can be identified on the Web in order to retrieve a representation of it.
2. **The Principle of Inconsistency**: A representation can be linked to URIs, and these links can be inconsistent (i.e., resulting in a URI that does not exist, returning something like the infamous “HTTP 404 Not Found” error).
3. **The Principle of Self-Description**: All representations should be linked to URIs that describe the meaning of the URI.
4. **The Principle of Least Power**: A representation should be described in the least powerful but adequate format.
5. **The Principle of the Open World**: The number of URIs, and so representations, can always increase.

These principles are a systemization that takes inspiration from the work of Berners-Lee and other designers of the Web (1998). Taken together, these principles are all independent. One can describe a possible web in which some combination of these principles does not hold. However, at this time we hold the conjunction of these principles minimally describe the philosophically interesting features of representations on the Web, and are so a good starting point to investigate the philosophical basis of the Web. Upon the surface, it does not appear that these principles may have philosophical weight, yet appearances can be deceiving. The connections between these principles and long-standing philosophical debates are productive and deep.

The first principle of universality posits that literally everything can be given a unique identifier, and as such relates to work on how identifiers like names are established. This is related to the debate between Kripke’s causal theory of names and neo-Fregean descriptivist accounts of names. It seems that the “meaning” of a web page on the Web can be given by the person who controls its URI, which is established in a clear “baptizing” process through domain name registration that is philosophically similar to Kripke’s theory of proper names. Alternatively, one could imagine that the “meaning” of a web page is given by the representations it returns in a neo-Fregean manner, so that even if the owner of the URI thought it was about blue cheese, if the URI returned pictures of the moon people would be correct in assuming that the particular URI was about a moon (Luntley 1999). Other principles point to areas that are under-explored by philosophy. The Principle of Self-Description demands a coherent story about what “self-description” even means. Self-description may be related to notions explored by Kolmogorov’s algorithmic information theory. The Principle of Least Power seems to demand a notion of power not easily mapped to traditional notions of power like Turing-completeness and the Chomsky Hierarchy.

The clearest correspondence to classical problems in artificial intelligence that external representations on the Web provokes is to problems of reasoning, but there are crucial differences like the Principle of Inconsistency. Unlike previous hypertext systems, on the Web any representation can be linked to any URI without a centralized database of links. In philosophical terms, if a link represents a logical predicate and a web page is a fact, then there is no principle of consistency on the Web. Long considered problematic for logic and philosophy, inconsistency is elevated to the status of a defining principle on the Web. Tolerance for inconsistency is precisely what removes the largest limiting factor of previous hypertext systems, since it allows users of the Web to link representations to URIs in whatever fashion they find most useful without asking permission. Combined with universality, this principle furthers the network effect, where the value of each representation grows in proportion to the size of the Web, since any representation may be linked to any other representation on the Web. In a historical note, it should be remembered that the original academic paper in which Berners-Lee attempted to present the idea of the Web was rejected precisely because of this inconsistency, yet it is inconsistency that allowed the Web to grow at such an astounding rate. Search engines like Google create post-hoc a centralized index of the Web, yet they are always behind the perpetual growth of the Web. For every moment Google crawls a web page, another web page can appear.

The Principle of the Open World states that the number of representations is always increasing. Therefore, unlike classical mathematics, it is difficult to ever say that a fact is false on the Web as a whole, since the set of representations one is reasoning about is ever-increasing. In order to pursue a strategy that says any particular fact is false or even unknown, one must somehow draw an arbitrary closed boundary over the Web, which violates the Web being an open system. Furthermore, since inconsistency is allowed, in an open system like the Web where decentralized agents are always adding new “facts” (new web pages being created, links being added, new logical
assertions made), it is untenable even with an arbitrary closed portion of the Web to say a fact is false, for facts can merely be inconsistent. This leads to another violation of traditional reasoning in classical artificial intelligence that tried to mimic human-reasoning by developing non-monotonic inference, because on the ever-increasing Web the results of an inference are just another part of the Web, so all inference should be monotonic. Due to this lack of a classical notion of true and false, the notion of truth on the Web can only be saved through intuitionism, in which truth is given by proof and any proof must take the form of a constructivist proof that does not rely on the Law of the Excluded Middle. So nothing is strictly false, “truth” is only what can be proven from a given set of facts selected from the Web and “false” that which cannot be proved from that particular set of facts. This exemplifies the insight of anti-realism of Dummett as applied to external representations; the debate between intuitionism and Platonism in mathematics has much wider philosophical repercussions (1959). On the Web anti-realism has engineering repercussions! This leads to an abandonment of the closed world solution to the Frame Problem in favor of a proof-theoretic notion of truth that can survive the open world of the Web.

Conclusions: The Semantic Web Reconsidered

By virtue of being an open system of external representations, the Web has a number of principles that directly conflict with classical artificial intelligence; yet, due to its heavy use of representations, the Web has naturally evolved a strange affinity with artificial intelligence. Search engines like Google occupy the space in popular imagination that the all-knowing robotic brains of classical artificial intelligence once did. The “Semantic Web” of Berners-Lee is at the heart of this fear of the return of AI. At the very first World Wide Web Conference, Berners-Lee announced plans to move the Web away from mere hypertext to a web of meaning.

To a computer, then, the web is a flat, boring world devoid of meaning...this is a pity, as in fact documents on the web describe real objects and imaginary concepts, and give particular relationships between them...adding semantics to the web involves two things: allowing documents which have information in machine-readable forms, and allowing links to be created with relationship values.

Long thought to be vanquished by the success of the neo-Heideggerian embodied mind, the spectre of classical artificial intelligence has seemingly returned on the Web in the form of the Semantic Web. The first step in this effort was the creation of a knowledge representation language for the Web. At the beginning the only research community involved heavily was the classical artificial intelligence community. Prominent champions of classical artificial intelligence such as Pat Hayes helped create its formal semantics (Hayes 2004). Yet, the Semantic Web is not classical artificial intelligence, it is in fact something new, as our principles above demonstrated. Classical solutions often do not work on the Web due to scalability issues and the open-ended nature of the Web. Due to these factors, formerly obscure areas of research like description logic, which guarantee decidability over open-ended data sets, are coming to the forefront of research in the Semantic Web. The revival of classical artificial intelligence on the Web makes perfect sense, since in the world of carefully engineered external representations, skills that humans lack but computers have in spades such as “logic-based reasoning or problem-solving in highly structured search spaces” can be crucial (Wheeler 2005). The type of problem the Semantic Web is meant to deal with is structured data-sharing and inference, which are more mundane than creating intelligence but perhaps just as useful in aiding human intelligence. Furthermore, with large reams of data, statistical methods often originating in artificial intelligence have even moreso than logic-based artificial intelligence proven to be crucial to the success of the Web, as search engines like Google demonstrate. Biological sensitivity makes little sense in the world of representations, for the question is not what can an intelligent human do, but how can computers complement an intelligent human. In the words of Andy Clark, we are “human-technology symbionts: thinking and reasoning systems whose minds and selves are spread across biological brain and non-biological circuitry” (2003). This symbiosis is done increasingly in practice today by the universalizing power of the Web.

Representations on the Web are machine-encoded, external, radically public, and reliant on a complex representational network of links. The meaning of these representations is ultimately grounded in their use by machines and humans. This is simply different from previous views of representations as primarily human-encoded, internal, private, governed by their truth conditions, and ultimately grounded in the neural activity of humans. While we would not dispute the possibility of the existence of internal representations, traditional artificial intelligence and analytic philosophy may be overemphasizing the role of internal representations in intelligence and underestimating the value of external representations. Equally so, the wildfire growth of the external representations as fueled by the Web should give us doubt about any anti-representationalist arguments about human intelligence. In fact, what the Web seems more concerned with is reducing what Floridi calls ontological friction: “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” (2007). It is precisely this minimization of ontological friction that the Web, and the Semantic Web, is trying to do. What is even more intriguing is the notion that at this point in time, non-biological representations may be more and more in the evolutionary driving seat. If this is indeed the case, then we have good reason to believe that the design and engineering of these external representations is no trivial task. Tim Berners-Lee has argued that “we are not analyzing a world, we are building it. We are not experimental philosophers, we are philosophical engineers.” One insight of the philosophy of the Web is that these representations on the Web do have philosophical significance, and philosophers are needed to help clarify their foundations. Another insight is that representations can indeed change the world, the Web being the example par excellence.

Endnotes

1. Originally the “Universal Resource Identifier,” now a Uniform Resource Identifier, such as http://www.example.org.
2. The most prominent being HTTP (Hypertext Transfer Protocol), although FTP (File Transfer Protocol) is also well known.
3. Such as HTML (Hypertext Markup Language) and XML (Extensible Markup Language).
4. See http://www.w3.org/Talks/WWW94Tim/ for the complete slides from Berners-Lee’s WWW 1994 announcement.
5. See http://www.xml.com/pub/a/2003/07/23/deviant.html for a synopsis of the argument between Hayes and Berners-Lee, as well as a wider critique.
6. I am in debt to all those, especially Wheeler and Clark, whose intellectual stimulation provided the foundation for this overview paper of my intellectual project. I am particularly thankful for the commentary received at the presentation of these ideas at the 2006 North American Computing and Philosophy Conference and at the presentation set up by Pat
Hayes at the Institute for Human and Machine Cognition. I am particularly in debt to Henry S. Thompson for encouraging my exploration of these ideas, and Peter Boltuc for encouraging me to contribute this paper and having much patience for my writing.

Works Cited


"In silico" Experiments: Towards a Computerized Epistemology

Jean-Gabriel Ganascia
affiliation?

Naturalization vs. Computerization

Recently, there have been attempts to naturalize epistemology with computers, i.e., to use information theory or computational models to explain how science and knowledge change (Dodig-Crnkovic 2007; Chaitin 2006). The proposition is that a general model of knowledge and of its evolution, which is what epistemology is all about, can be grounded on digital computation and information theory. The underlying idea is not new and comes from Leibniz and Hobbes, to name but two. The principle on which all these endeavors are based is that we really understand something if we are able to compute it. Since Alan Turing’s famous article on computable numbers (Turing 1936), we know that all computations can be executed on programmable digital machines. Therefore, one of the consequences of the above principle is that we understand something if we can program it on digital machines. Thus, the attempts to naturalize epistemology can be assimilated to efforts to build a digital epistemology (Chaitin 2006).

Our present goal is more modest: it is not to contribute directly to an informational or computational naturalization of epistemology, but to start from the observation that today computers are everywhere, and this transforms both scientific practices and epistemology. In other words, the way we understand the world is greatly influenced by the general use of digital computers, which is not the same as saying that understanding is computing. Since all facts are now reduced to information sets and recorded on huge memory devices, it is possible to directly test theories on recorded data without having to carry out experiments evaluating all of them in the outside world. Undoubtedly, this changes scientific activity, at least in part, and many real world experiments no longer need to be performed, which seems highly desirable for economic and ecological reasons. Consequently, the general use of in silico experiments will have to be promoted. But, the exact status of these in silico experiments and the validity of the knowledge the scientists infer are open questions of epistemology. In a way, in silico experiments are similar to “thought experiments” (Mach 1976; Sorensen 1992; Dennett 1995; Ganascia 2002): both types are achieved outside their object world, in a virtual world which will be mental for thought experiments, and digital for in silico experiments.

In the past, many philosophers—including Karl Popper—have criticized the role played by “thought experiments” in science (Popper 1959), saying that they did not provide any strong scientific justifications. Some of these philosophers considered these experiments could even be misleading if they picture scientific concepts incorrectly for pedagogical purposes. For instance, Karl Popper criticized what he called the misuse of “thought experiments” in quantum physics (Popper 1959).

Can similar criticisms be leveled at in silico experiments? One of the goals of this article is to answer this question. To be more precise, epistemology, understood as the branch of philosophy that studies knowledge and the way knowledge is acquired, has to take into consideration new scientific procedures that make extensive use of computers. The aim here is to clarify their status and to show that the evolution of epistemology subsequent to the computerization of science will open up new philosophical perspectives.

Actual and Virtual Experiments

Science and the Senses

In ancient times, science was first and foremost a question of observation, and for Plato the most important sense was that of sight. Later on, in modern times, touch took over from sight: people wishing to understand the natural world spent more and more time provoking the subjects they were studying. Thus, in the sixteenth century, Andreas Vesalius (1514-1564) renewed human anatomy by dissecting the corpses of people condemned to death. Scientific experimentation, in its modern meaning, corresponds to this reversal: it is not enough just to observe, and a scientist will intervene on the world in order first to understand it and then to transform it. This active intervention on the real world continued relentlessly: soon, autopsies no longer satisfied naturalists, who chose to provoke natural phenomena on the living body in order to understand the life
springs. They then went further and started performing what are known as in vivo experiments because they are carried out on living beings. And so it went on: investigation was not only a question of touching and provoking nature, but also of reconstructing it. This led to the idea of reproducing in vitro, i.e., in glass test-tubes, the chemical reactions that are at the origin of the elementary physiological functions.

In the eighteenth century, some scientists, including Jacques de Vaucanson (1709-1782), created artificial physiologies to get a better understanding of animal functions. Those automata imitated living beings by the means of mechanical devices. Computers are but an extension of this trend: digital data processing techniques can henceforth mimic almost all natural mechanisms, in particular those of living beings, by reducing them to informational processes. This has given birth to a new type of experiment, which no longer has recourse to the external senses but merely unravels sequences of logical operations. As already pointed out, these experiments are said to be in silico because they are performed neither on living beings nor on living matter, but on silicon chips which execute logical operations of data processing. In that in silico experiments take place virtually, without actually touching the subject under investigation but only altering their representations, they look like virtual experiments, similar to “thought experiments.” However, to clarify the actual status of in silico experiments and to show how different they are from “thought” experiments and in vivo and in vitro ones, it would be useful to recall the epistemological status of classical experiments.

Claude Bernard's Closed-Loop Discovery

Claude Bernard (1813-1878) was one of the most eminent nineteenth-century physiologists, and a pioneer in many respects. He introduced the concept of internal environment (the “Milieu intérieur”) (Grmek, 1997), which corresponds to today’s principle of “homeostasis.” He investigated and explained many physiological mechanisms, including the glycogenic liver function (Prochiantz 1990), the effects of carbon monoxide (Bernard 1864; Grmek 1973), and the effects of curare (Bernard 1857; Bernard 1864). But Claude Bernard was not only a great physiologist, he was also a theoretician who generalized his experimental method in his famous book Experimental Medicine (1927), which is today a classic that all young medical students are expected to have read.

According to his views, scientific investigation can be reduced neither to the sole observation of facts nor to the construction of theories that have not been previously confirmed by empirical evidence. In other words, Claude Bernard is not an inductivist, who reduces scientific activity to the pure induction of general rules from the observation of particulars, nor is he an idealist—or a neo-Platonist—who thinks that ideal, pure, and perfect theories are found before any experiments are carried out. The experimental method he promotes begins with an initial theory, which is usually built from passive observations or preconceived ideas. When the phenomenon is unknown, some experiments are performed just “to see what happens.” Claude Bernard does not explain how the first idea or the initial theory is put forward, just before a concrete experiment, and its role is twofold. First, it tests the verisimilitude of the hypothesis, i.e., it shows whether or not the hypothesis under investigation is contrary to common sense and our past experience. Second, it helps to design experiments with respect to the current hypotheses, i.e., to build physical devices which will generate the observations that will validate or invalidate a theory. However, it may well happen that, for technical, ecological, or ethical reasons, an experiment is impossible, in which case a thought experiment replaces a real world experiment and helps to evaluate the consequences of a theory. As we shall see in the following, to each of these thought experiments correspond some in silico experiments that have a similar place in the discovery cycle even if, due to their computational nature, the role they play is different.

Once an initial theory is given, the scientist must design an experimental apparatus able to test (corroborate or refute) it. The experiments are viewed as “provoked” observations generated by an appropriate device, and these observations are compared with the expectations derived from the given initial (current) theory. A careful analysis helps to revise, correct, refine, or validate the current theory, after which a new experiment is devised to validate the refined theory, and the experimental method is iterated until the theory predicts all current experimental results. In a way, this experimental method is cyclic, which is the reason why it is known as a “closed-loop discovery” process.

To be more precise, the experimental method described by Claude Bernard is an iterative procedure of theory refinement that proceeds in three steps, each step involving a specific scientific function:

**Experimentation:** a hypothesis that has to be validated has to be given. It is called an idea or a theory. For the sake of clarity, we shall refer to it as the current theory. The first step then is to design an experimental apparatus able to generate observations that can be compared to expectations derived from the current theory. In other words, the experiment is designed to test the hypothesis under investigation, i.e., the current theory.

**Observation:** the second step involves collecting observations from the designed experiments. It is not only a passive step, since the experimenter may interpret observations and note unexpected details.

**Analysis:** this third step is the most crucial and original. The current theory predictions are compared to the observations and, where necessary, i.e., when the predictions do not match the experimental observations, new plausible hypotheses are generated to transform the current theory.

Some artificial intelligence research aiming at the computational reconstruction of scientific discovery (Kokabas & Langley 1998) has focused on a very similar cyclic discovery process. Recent attempts to automate scientific discovery using a “robot scientist” which could generate and test hypotheses by itself also refer to the notion of closed-loop discovery (King et al. 2004). So, this discovery cycle still seems to be valid.

**Existence and Role of Mental Experiments**

Given this general description of the experimental method, two questions are of interest here. The first is about the existence of virtual experiments in this cyclic discovery process, the second concerns the role they play in the discovery loop.

In the case of Claude Bernard, the answer to both questions is easy: many of his notebooks, for instance the red notebook (Bernard 1942), contain suggestions for experiments including some that were actually carried out and reported in the laboratory notebooks. These suggestions can be seen as particular cases of mental experiments, which are required preliminaries to any factual experiments. More generally, Ernst Mach argued that a thought experiment was “a necessary precondition for physical experiment” (Mach 1976). Therefore, a thought experiment takes place once a hypothesis has been put forward, just before a concrete experiment, and its role is twofold. First, it tests the verisimilitude of the hypothesis, i.e., it helps to design experiments with respect to the current hypotheses, i.e., to build physical devices which will generate the observations that will validate or invalidate a theory. However, it may well happen that, for technical, ecological, or ethical reasons, an experiment is impossible, in which case a thought experiment replaces a real world experiment and helps to evaluate the consequences of a theory. As we shall see in the following, to each of these thought experiments correspond some in silico experiments that have a similar place in the discovery cycle even if, due to their computational nature, the role they play is different.
others facilitate communication between the scientists and the public, or among researchers. Since the goal here is not to discuss mental experiments for their own sake, but to look at \textit{in silico} experiments, this last type will not be considered any further, as it does not seem to have any equivalent in the world of computer-aided experiments.

\textbf{In Silico Experiments}

As we have seen, thought experiments have three main functions in the discovery cycle. The first is to evaluate the verisimilitude of a theory with respect to our past experience; the second is to help design experiments by anticipating the consequences of the conflicting hypotheses; the third is to determine the consequence of the theory when real experiments are not possible. As said above, for each one we find a class of \textit{in silico} experiment. Let us now be more precise.

\textbf{Informational Experiments}

For more than fifteen years now, data-mining techniques have been used to test hypotheses against massive quantities of digital information, their explicit goal being to automatically discover knowledge from databases (Klosgen et al. 2002). They are based on artificial intelligence techniques that produce many tentative hypotheses and then evaluate each of them with respect to digital data. In a way, these hypothesis evaluations can be seen as experiments, since each of them compares the consequences of a hypothesis to data contained in databases. Because these experiments are done on computers, they are typical cases of \textit{in silico} experiments.

Practical applications of such informational experiments are many and varied. Insurance companies, for instance, use such techniques to help assess individual risks. In scientific activities, it has become necessary to devise systematic informational experiments when huge quantities of data are generated by captors because there is no other way to give sense to these data. For example, in molecular biology, the linear structure of macromolecules such as DNA or proteins is now determined more or less automatically; the results are stored in huge databases, and computers are required to interpret these enormous amounts of information (Danchin et al. 1991).

From an epistemological point of view, informational \textit{in silico} experiments reverse the traditional conception of experimentation. In classical approaches such as Claude Bernard’s discovery cycle, the experiment is seen as an active intervention on the natural world and is designed with respect to the theoretical hypothesis it has to validate or invalidate. In a nutshell, first there is the hypothesis, then an experimental apparatus to validate the hypothesis is designed, and lastly data are generated. In other words, empirical data come afterwards. In the case of informational \textit{in silico} experiments, it is the data that come first, after which the hypotheses are generated and tested against the data. The data are collected before the hypothesis has been put forward. In a way, the logic of experimentation has been reversed by the generalized use of computers, which opens up new avenues in science. In medicine, for instance, this could transform clinical evaluations, which would able to use all existing patient databases.

\textbf{Simulations}

The second and the third types of thought experiments mentioned above consist in evaluating the consequences of theories, either to facilitate the design of real world experiments or to avoid actually carrying them out. Thanks to their simulation abilities, computers can play a similar role and they may anticipate the consequences of theories. What is more, computational simulations are far more accurate than mental experiments because they can be quantified precisely and compared with empirical evidence. Therefore, they may help reduce considerably the number of real world experiments, which is desirable not only for economic reasons but also for ecological and ethical ones. This already happens in many domains; for instance, the testing of nuclear weapons has been greatly reduced thanks to computer simulations, and much progress has been made in the physics of materials, in climatology, in the environmental sciences, etc., using computer models that anticipate natural phenomena.

Computational simulations integrate theories and we could say that the computational models on which simulations are based are theories. Thus, by specifying the input conditions and by running the simulation program, an experiment is performed in a virtual world that represents a theory with operations on symbols. It is clearly an \textit{in silico} experiment.

All these simulations involve the transformation of representation and this last point needs to be underlined: \textit{in silico} experiments presuppose some explicit and well-defined symbolic representation on which calculations operate. These symbols can be restricted to “sub-symbolic” features, i.e., numbers, but in all cases computations transform symbols that represent reality. In technical terms, symbolic representations rely on “ontologies” that associate inference mechanisms to sets of features. As they are only representations, these computational models are approximations of what they intend to represent. It follows that the goal of simulations is not necessarily to reduce reality to a calculation, but to anticipate some aspects of the reality through the use of a model.

\textbf{Info-Computational Models}

As said earlier, \textit{in silico} experiments essentially encompass two principles, one which corresponds to an informational approach, i.e., to the evaluation of a hypothesis on prerecorded data, and the other to a computational view where theories are formalized and simulated using computers. Note that, as mentioned in Dodig-Crnkovic (2007), two philosophical views, informationalism and computationalism, tend to favor reducing everything either to information, which is considered as the substance of the universe, or to calculations on which all changes, both physical and mental, are supposed to be based. This could mean that some \textit{in silico} experiments resort to informationalism, others to computationalism, which would justify an info-computationalist synthesis as proposed by Gordana Dodig-Crnkovic (Dodig-Crnkovic 2007). However, in the case of \textit{in silico} experiments, both views are approximations. Databases do not contain exhaustive information on the world, but only a partial view of it, which means that hypotheses validated using informational experiments have to be confirmed. In the same way, digital models on which computer simulations are based are only partial approximations of reality (Noble 2006); their scope is always restricted and there is no absolute world computer model. As a consequence, our goal is not to reduce epistemology, i.e., the theory of knowledge, to an informational view, to a computational perspective, or to a combination of the two, but to see how computers in general, and \textit{in silico} experiments in particular, modify the activity of scientists and the production of knowledge.

\textbf{Computerized Epistemology}

\textbf{The C\textsc{y}ber\textsc{n}ard Project}

Epistemology has to take into consideration the way computers transform scientific activity and, in particular, experimental validation procedures. But epistemology may also be transformed by the use of computers in the reconstruction of old scientific discoveries. For example, work has been done on the rational reconstruction of old scientific discoveries using computers (Langley et al. 1987). However, this work does not really question the status of experimentation; most of it is based

— 13 —
on past observations and tends to reduce scientific discovery to an inductive step. The goal of the CYBERNARD project is to understand Claude Bernard’s experimental route (Ganascia & Debru 2007) and to show that in silico experiments can reproduce the suggestions for experiments reported in Claude Bernard’s notebooks. Before each suggested experiment, there are some conflicting hypotheses that need to be discriminated empirically. The simulation of these suggestions for experiments, under different theoretical hypotheses, shows the logical function of each of the planned experiments. It is then possible to see how theoretical investigations evolve with time. What this article claims is that such an attempt could profoundly modify methods in epistemology, since it makes it possible to reconstruct the representations scientists make in their everyday activities.

Core Models
To design a computational model that simulates the intellectual route leading Claude Bernard to his discovery, we have supposed that he had in mind what we call “core models” that contain basic physiological concepts—such as internal environment, organ names, etc.—upon which he built his theories.

More precisely, Claude Bernard had in mind an ontology, which was explicitly described in his writings: he presumed that organisms are composed of organs, themselves analogous to organisms since each of them has its own aliments, poisons, stimulations, actions, etc. The internal environment, mainly the blood, carries organ poisons and aliments, while the actions of the organs may have different effects on other organs and, consequently, on the whole organism.

Theories correspond to hypothetical organ functions that Claude Bernard wanted to understand and explain, while “core models” describe the physical architecture of the organism reductions that were necessary for Claude Bernard’s conceptualization. These “core models” constitute the core on which the reasoning process is based and, depending on the question under investigation, may be more or less simplified. For instance, if one wants to investigate the function of the heart, it is not necessary to detail the precise role of all muscles. The CYBERNARD project builds and simulates these “core models” using multi-agent architectures. Such simulations have to give a simplified view of Claude Bernard’s representation of both the normal behavior of the organism and the consequences of a dysfunctions organ due, for instance, to some toxic substance.

A second level of the model manages hypotheses relative to the function of different organs. Each working hypothesis is evaluated through empirical experiments. Claude Bernard assumed that one can use toxic substances as tools of investigation—he suggested the idea of a “chemical scalpel” —to dissociate and identify the functions of different organs. He presupposed, as an underlying principle, that each toxic substance neutralizes one organ at a time. When a toxic substance affects an organ, the anatomy of death shows how the organism behaves without the poisoned organ. Nevertheless, even when working on such a presupposition, many physiologists are puzzled by the investigation because it is a double entry enigma: they have to understand and explain both the organ(s) contaminated by toxic substances and the function of the affected organ(s).

Our aim is to use “core models” to simulate all suggestions for experiments and to understand the place of these experiments in the discovery process using the hypothesis management module.

Towards an Electronic Epistemology
The description of in silico experiments given in this article shows how computers modify scientific activity and that, in a way, in silico experiments are instances of thought experiments. It is also argued that the criticisms leveled against thought experiments by philosophers, especially by Karl Popper, do not target in silico experiments. Although this point has not been fully dealt with here, it is certainly the role of contemporary epistemology to discuss such questions because this concerns the present state of the sciences and the way knowledge is produced today. And as the description of the CYBERNARD project illustrates, the practice of epistemology may be transformed by the introduction of computer models.

To sum up, the extension of the place of computers in everyday life has an impact on epistemology because the subjects studied and the methods used are changing. These changes are not only in silico and are not limited to reducing epistemology to an info-computational model. Modern epistemology is not condemned to become an abstract theory of knowledge and of how it is built, and may continue to consider the details of scientific activities, with concrete references to historical developments. To use another Latin phrase, this new computerized epistemology can be said to be in situ, since it is located in the concrete context of discoveries. To quote Gaston Bachelard,

It will be a phenomenology of the studious man, of the man tense in his study and not only a vague assessment of general ideas and acquired results. It will have to make us attend the daily drama of daily study, to describe the rivalry and the cooperation of the theoretical effort and experimental research, to place us at the center of this perpetual conflict of methods, which is the manifest character, the tonic character of the contemporaneous cultural science. 2

However, this modern epistemology, which takes into account the influence of computers on daily scientific activity, is not restricted to the naturalization of epistemology, i.e., to the reduction of epistemology to a “science of nature.” The cross-road between epistemology and computer science offers many alternatives. Some of them have been presented here, and they could be considered as anti-naturalizations. If we take the opposition between the “science of nature” and the “science of culture” that was introduced by the German philosopher Heinrich Rickert (1863-1936) we could say that computerized epistemology is not only a subject of the science of nature; it is also a science of culture, which shows and measures the influence of computers on the production of knowledge today.

Appendix: Making a Brief Latin Revival
It would seem that it is thanks to computers that Latin, a so-called dead language, is today enjoying a brief revival. At the end of the eighties (cf. http://en.wikipedia.org/wiki/In_silico), biologists who wanted to give a name to experiments performed using computers or, to be more precise, using the silicon microchips that constitute the core of computers invented a new Latin idiom, in silico. The expression was built by analogy from—and in contrast to—in vitro experiments, i.e., experiments on living organisms, and in vitro experiments that bind biological mechanisms to chemical processes reproduced in glass test-tubes.

Some purists (Quinion 2006) seem to be unhappy with this new idiom because, strictly speaking, linguistic rules which could have been emulated on a computer would have given the neologism in silicio from the Latin word silicium, which means silicon in English. Nevertheless, usage has imposed the idiom...
in silico, which is widely accepted today. Although computers led to a brief Latin revival through the invention of a new Latin idiom, this was, however, due neither to a computer simulation nor to an informational operation, but to a new formulation of old questions because of the generalized use of digital processors. In a way, we find the same thing in contemporary epistemology: with computerization this discipline evolves, but this is due not only to an informational or to a computational model of epistemology, but also to the increasing role that computers play in science in general, and to the increasing role they could play in epistemology in particular. This article constitutes an attempt to highlight this role.

Endnotes
1. See the Appendix to this article.

References

COMMENTARY ON HARMAN

Verstehen and the Explanatory Gap
Sara Worley
Bowling Green State University

Harman suggests in his provocative piece that we can come to some understanding of the “explanatory gap” by considering the phenomenon of verstehen. Verstehen is the methodology recommended by some writers for coming to understand social phenomena. The idea is that we can’t adequately understand these phenomena by arriving at purely objective descriptions of them, even if such descriptions give us some predictive power. Instead, to fully understand them, we have to understand what their meaning is for their participants, and that requires discerning the participant’s point of view. The parallel with the explanatory gap is clear. Nagel, for example, argued that we can’t know what it’s like to be a bat, no matter how much knowledge we have about bat neurophysiology and echolocation. Acquiring objective knowledge doesn’t tell us what it’s like from the first person point of view.

Harman also suggests, however, that we can understand failures of verstehen by considering the difficulties of translation.

In order to interpret a language, one has to have the appropriate concepts in one’s own language. One can’t translate a term into one’s own language if one doesn’t have the appropriate concepts. The reason we can’t understand what it’s like to be a bat is because we don’t have the appropriate concepts in our own language in order to do the translation.

But of course none of this has any untoward metaphysical implications. We may not be able to tell whether something is a dance or a religious ritual if we have a description of it. To know which it is, we need to know how the participants experience it, how they understand it from their point of view. But there isn’t a metaphysical problem about the existence of dances or religious rituals. Nor is their existence any challenge to physicalism. Just so, we shouldn’t think that the existence of the explanatory gap, or the fact that some things are only accessible from the first person, poses any challenge to physicalism.

However, it doesn’t seem that this appeal to verstehen should be very satisfying to those who think that the explanatory gap arguments do raise serious metaphysical problems. The problem raised by the explanatory gap is not simply the problem of different theoretical frameworks. The problem, rather, is why
some frameworks can be acquired by processes of learning and acculturation, and others can not.

The point can be illustrated by considering the problem of diagnosis in psychiatry. There are a number of different theoretical frameworks for thinking about mental disorder, including, inter alia, psychoanalysis, cognitive behavior theory, interpersonal theory, and biological psychiatry. Given a particular pattern of behavior, each of these different theoretical frameworks may recommend a different diagnosis, etiology, prognosis, and treatment plan. But none of these theoretical or conceptual differences involve anything analogous to the explanatory gap. The key is that, although there are conceptual differences between each of these theories, still, each is available objectively, from the third person. Anyone who takes an interest can learn any of these theories, and learn to classify the world accordingly. These are different theoretical frameworks, but they are all equally third person or objective, in Nagel’s sense.

Now, of course, one might argue that what one is capable of learning depends on what capacities and background one has. One can’t very well expect someone from a pre-modern culture to pick up a textbook of biological psychiatry and make much headway with it. The textbook will involve too much background knowledge and assumptions which he will not share, at least not without a lot of preliminary training.

Now, of course, the same is true in reverse. The modern western anthropologist who arrives in a new culture might not immediately be able to understand the culture from the “inside.” He or she might not have the right concepts, so may not be able to grasp the theoretical framework, the ways of understanding their experience and their world, used by the participants themselves. But this anthropologist can presumably acquire the relevant concepts by becoming immersed in the culture and undergoing the appropriate “learning experiences.” This must be possible because, after all, children are not born understanding their culture, or the theoretical frameworks inherent to it. They acquire these things, via being immersed in the culture, learning the language, being corrected and reinforced, etc.

The person originating in a pre-modern culture, then, and the western anthropologist, then, can each acquire, via appropriate training and experience, the concepts and theoretical frameworks appropriate to the other’s culture. They can acquire these things because language and concept learning essentially involves responses to publicly available, “objective” stimuli. Otherwise, it’s hard to see how it would be possible for children to learn language, or to become acculturated. So each can learn to see the world through the theoretical frameworks used by the other, and, thus, learn to see the world from the other’s point of view.

But how is all this supposed to enlighten us about the explanatory gap? The western anthropologist can learn what it’s like to be a member of a particular culture, can learn the theoretical understandings that they have of their practices and experiences, by being appropriately immersed in that culture. The member of a pre-modern culture can similarly acquire western scientific understandings by being appropriately immersed in a western scientific culture. But can we learn what it’s like to be a bat by living with bats? Is what’s missing simply a theoretical framework, which can be acquired as human cultures are acquired, by a process of learning and acculturation? Those sympathetic to the explanatory gap would presumably answer this question in the negative. Acculturation and learning are not what’s missing. What’s missing is the peculiarly first person phenomenal experience, and this can’t be acquired simply by the sorts of learning involved in acquiring a new theoretical framework, or a new language.

The proponents of the explanatory gap may thus agree with Harman that the problem is that we don’t have the right concepts, so that we can’t understand what it’s like to be a bat. But the problem then becomes that some concepts (the objective, third personal ones) are available to anyone with appropriate exposure and acculturation, and others are not. It’s the status of these peculiarly first personal phenomenal concepts that Nagel finds mysterious, and the appeal to verstehen and failure of translation does not help resolve this mystery.

----

Commentaries on Floridi

Information Ethics as Macroethics: Perspectives for Further Research

Antonino Vaccaro
Carnegie Mellon University & CEBE–FCEE Catholic University of Lisbon

“Understanding Information Ethics” is a very interesting synthesis and critical review of a long series of studies published by Luciano Floridi between 1999 and the current date (i.e., 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 would start this commentary pointing out that I find very interesting and useful how Luciano introduces and explains the genesis and the fundamentals of Information Ethics, and its related technological and sociological scenarios. Moreover, this paper clearly demonstrates two uncommon capabilities of its author, i.e., the capacity of clear and comprehensive synthesis and the ability to provide bold and reliable forecasts for the evolution of digital societies.

The main characters of this narration are very well known to the reader. Information and Communication Technologies (ICTs) are dramatically changing how individuals and organizations behave and inter-act. But, while ICTs are offering several opportunities for the well being of people, they are also the cause of social problems and moral dilemmas which point out “profound philosophical questions about human nature [and] the organization of a fair society.”

The conceptual apparatus adopted in Floridi’s discourse is based upon two main concepts, i.e., the “infosphere” and “re-ontologization.” They are used to draw a simple but extremely elegant Theory of Information Ethics and to define its main characters as microethics first and lately as a macroethics.

My commentary will focus on just some possible perspectives for further research offered by Luciano’s works and in particular by the Theory of Information Ethics (IE) as a Macroethics.

One of the challenges of the future is analyzing what kinds of answers the Theory of IE as Macroethics is able to offer in day-to-day life to human beings that have the power to create, manipulate, and destroy qua information objects and to create autonomous agents which in turn have analogous or even more powerful capabilities in the infosphere (the creation, manipulation, and destruction of higher amounts of objects—qua information). Two main ways (or a combination of both) can be followed to analyze this issue. They are related to two different novel contributions of the Floridi’s Theory.
One of the original points of the IE Theory is to enlarge the horizon of entities which have the right to be included in the analysis of ethical issues. Whether Bio-centric Ethics includes all constituents of the living world and Land Ethics considers all animated and un-animated things, the IE as macroethics embraces all informational entities. A point that should be addressed in the future is the impact made by the inclusion of informational entities to the analysis of applied problems. In order to approach this issue, further research should look at: a) practical situations in which traditional entities (i.e., those entities analyzed in previous macroethics theories such as human beings) and new informational entities (i.e., information entities that were not included in previous macroethics theories such pure cyber entities) interact, and b) practical situations in which different kinds of new informational actors interact between themselves.

The Floridi Ethical theory offers another interesting innovative contribution. Given the new ethical domain, i.e., all informational entities, it provides new ethical principles. In particular, looking at the relation human beings-infosphere, I believe it could be very interesting to understand whether the ontological/equality principle, which defines Spinozian rights over all informational objects, could provide new normative indications to the concept of ethical right and good in human behavior. The analysis of this issue in different real contexts of the day-to-day life of individuals and organizations can provide useful indications of the robustness and completeness of the Floridi’s theory which sometimes could appear overly general. The analysis of the ontological/equality principle can also be used in the resolution of practical cases involving artificial intelligence and robotic engineering. Indeed, what are the limits and the norms we should impose to informational artificial autonomous agents?

I would like to point out another possible direction for further research. As already noted, the IE Theory represents an extension of the Land Ethics by including all informational entities and by offering new ethical principles. I think it could be very interesting to understand what happens if we apply the ethical principles of the IE (e.g., ontological/equality principle) to situations in the material (in opposition to the digital) world. In other terms, it could be very useful to test whether the IE Theory and, in particular, its new ethical principles are able to provide different answers and perspectives to problems which were traditionally in the domain of older macroethics such as Bio-centric or Land Ethics.

References

So You Say You Want a Re-Ontological Revolution: Comments of Luciano Floridi’s “Understanding Information Ethics”

John P. Sullins
Sonoma State University

Reading Luciano Floridi’s paper (2007) caused me to recall the time I destroyed a world. My daughter had been playing a computer game called Animal Crossing, which involves creating a simple avatar child that one controls as it wanders around a small virtual world populated with cartoonish AI bots. There is no real point or objective to the game, one just encounters the software agents in the game world and your character eventually ends up doing things like working at a job, paying a mortgage, doing errands, writing letters to other software agents, picking weeds, decorating one’s house, etc. My oldest daughter had been playing this for a few months and had built a teaming virtual world. One day while trying to help her save the game, through a combination of my ineptitude and a glitch on the memory card of our Nintendo Game Cube, I irrevocably lost all the saved data from the game. There was no going back, the game syncs with the time and date of the real world and randomly generates the hundreds of details that go into each of the worlds making each player’s experience different and each game world a unique creation. My daughter was crushed, all her AI friends were gone, her virtual world destroyed, and I its agent of destruction. Of course, this event is nothing in the grand scheme of the universe, but as more and more of the meaningful activities of our lives are moving to the digital world, it is likely more serious situations analogous to this one will occur. Information Ethics is providing us with tools and methods that will allow us to sensibly speak about the moral obligations, both trivial and robust, that we have towards the information objects, systems, and entities that already impact our lives.

Floridi claims that “...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” (Floridi 2007, 4). Global ICT and web technologies are converging into what Floridi calls an infosphere making us all,
like my daughter and the others who build and interact with virtual worlds, denizens and designers of a new digital ecology. Floridi argues that these technologies are not simply interfering with the ecology we live in; instead, they re-ontologize our ecology by re-engineering its intrinsic nature. Floridi argues that this transformation is so new and pervasive that it stands to, at the very least, transform the study of ethics and morality if not the entirety of the philosophical endeavor. This move turns philosophy into the deep questioning of the nature, and ethical use of information, and will provide us with powerful new ways to approach long-standing problems in moral philosophy, methods that may prove useful in the other subjects of philosophy as well.

This is where Floridi’s theories part ways with other thinkers who have postulated the re-ontologizing effects of technology such as Heidegger (1954; 1977), Borgman (1984), and Baldrillard (1981; 1994). All of these authors argue that the re-ontologizing forces of technology are potentially dangerous because they move us away from the more or less authentic world and into a simulacral world constructed out of high technology. Floridi’s theory lacks this melodramatic narrative; for him, the re-ontologization does not involve the transforming of the intrinsic nature of the natural world. Nature is not changed into information technology; rather, in his theory we see the ontic structure of our ecology as informational and information technology as a proper tool for its study. This is a liberating theory for the endeavor of philosophy as it allows us to formulate an information ethics that utilizes computational tools and techniques.

In this way, information technology is not occluding our view of reality, but instead giving us a way to bring it into sharper focus. Indeed, this is a very powerful new way to approach problems in technoethics and moral theory in general. I have argued that the theory of information ethics allows us to make much more sense of the nature of artificial moral agency allowing us to extend moral status well beyond just humans and other robust cognitive entities (Sullins 2006 and 2008).

There are many things that need to be said about Floridi’s theory but I will limit myself to only one question here. One of the most exciting aspects of information ethics is that it allows us to fully realize the power and challenge of the deep ecological worldview and extend it to new domains. Arguing sensibly for the moral rights of animals, environments, and rocks is increasingly more difficult as you climb down the great chain of being that still shackles our imagination. Traditional ethics is increasingly more difficult as you climb down the great chain of being that still shackles our imagination. Traditional ethics is far too anthropocentric, causing its practitioners to fail to see that ecosystems can have any moral rights separate from the humans living in them. Seeing the natural world instead of being that still shackles our imagination. Traditional ethics is increasingly more difficult as you climb down the great chain of being that still shackles our imagination. Traditional ethics is far too anthropocentric, causing its practitioners to fail to see that ecosystems can have any moral rights separate from the humans living in them. Seeing the natural world instead of being that still shackles our imagination. Traditional ethics is increasingly more difficult as you climb down the great chain of being that still shackles our imagination. Traditional ethics is far too anthropocentric, causing its practitioners to fail to see that ecosystems can have any moral rights separate from the humans living in them. Seeing the natural world instead of being that still shackles our imagination. Traditional ethics is increasingly more difficult as you climb down the great chain of being that still shackles our imagination. Traditional ethics is far too anthropocentric, causing its practitioners to fail to see that ecosystems can have any moral rights separate from the humans living in them. Seeing the natural world instead of being that still shackles our imagination. Traditional ethics is increasingly more difficult as you climb down the great chain of being that still shackles our imagination. Traditional ethics is far too anthropocentric, causing its practitioners to fail to see that ecosystems can have any moral rights separate from the humans living in them. Seeing the natural world instead of being that still shackles our imagination. Traditional ethics is increasingly more difficult as you climb down the great chain of being that still shackles our imagination. Traditional ethics is far too anthropocentric, causing its practitioners to fail to see that ecosystems can have any moral rights separate from the humans living in them. Seeing the natural world instead of being that still shackles our imagination. Traditional ethics is increasingly more difficult as you climb down the great chain of being that still shackles our imagination. Traditional ethics is far too anthropocentric, causing its practitioners to fail to see that ecosystems can have any moral rights separate from the humans living in them. Seeing the natural world instead of being that still shackles our imagination. Traditional ethics is increasingly more difficult as you climb down the great chain of being that still shackles our imagination. Traditional ethics is far too anthropocentric, causing its practitioners to fail to see that ecosystems can have any moral rights separate from the humans living in them. Seeing the natural world instead of being that still shackles our imagination. Traditional ethics is increasingly more difficult as you climb down the great chain of being that still shackles our imagination. Traditional ethics is far too anthropocentric, causing its practitioners to fail to see that ecosystems can have any moral rights separate from the humans living in them. Seeing the natural world instead of being that still shackles our imagination. Traditional ethics is increasingly more difficult as you climb down the great chain of being that still shackles our imagination. Traditional ethics is far too anthropocentric, causing its practitioners to fail to see that ecosystems can have any moral rights separate from the humans living in them. Seeing the natural world instead of being that still shackles our imagination. Traditional ethics is increasingly more difficult as you climb down the great chain of being that still shackles our imagination. Traditional ethics is far too anthropocentric, causing its practitioners to fail to see that ecosystems can have any moral rights separate from the humans living in them.

Still, as the frictionless infosphere commingles with the natural world and the two become interactive and dependant on one another, we may enter a new renaissance of domestic bliss. When my daughter’s weed picking activities online result in actual weeds disappearing in our back yard, we will become a very happy family indeed. Let’s just hope the infosphere we are now constructing has a proper backup system.

References


Sullins, John P. “When is a Robot a Moral Agent?” *International Review of Information Ethics* 6 (December 2006). www.i-r-i-e.net.


**Discursive Explorations in Information Ethics**

Elizabeth A. Buchanan

*University of Wisconsin–Milwaukee*

In “Understanding Information Ethics,” Luciano Floridi (2007) issues a call to arms for all citizens of the digital world, the infosphere, as it is termed, a realm of epistemological and ontological synergy. We are called, in no short terms, to reconsider our ways of knowing and being as members of this sphere. It is unarguable, and perhaps inevitable, that our realities are now deeply and complexly intertwined. The line between the real and the virtual, online and onsite, have blurred beyond recognition. It does, in fact, sound anachronistic to speak in such terms, so enmeshed we are in the infosphere.

The infosphere, in Floridi’s view, is a net, with open, fluid peripheries. It is unlike any physical counterpart we have thusly known; and, it is not an entity to which we can enter and leave freely. We are, as Henry Jenkins has artfully described, living in a “convergence culture.” Once “in,” one does not, can not, leave. Indeed, David Noble (1998) talked about the incessant technological umbrella in which we as faculty now work and its “mandatory” connectivity, illustrating its negative impact on and implications ten years ago. And yet, here we stand more firmly embedded than ever within that inexhaustible state of connectedness. Such transparent connectivity is both a blessing and a curse, in Floridi’s theorizing. We are at once master and slave, controlling and controlled, depending and dependent. This constant duality has significant consequences, as Floridi intimates. But, what emerges, so clearly and dramatically, out of “Understanding Information Ethics” is an explicit illustration of the growing responsibilities we as professionals and as citizens must recognize and address. An important question is, of course, How do we fulfill these responsibilities?

To set the stage for this question, Floridi categorizes the traditional and emerging information ethics issues through the conceptual frameworks of:

1. Information-as-a-resource ethics (availability, accessibility, accuracy, and their related outcomes, digital divide, infoglut, and reliability and trustworthiness of information);

2. Information-as-a-product ethics (accountability, liability, plagiarism, etc.);
3. Information-as-a-target ethics (hacking, censorship, freedom of expression, intellectual property contestations, etc.).

Following this three-fold distinction, Floridi articulates the shortcomings of such microethics models. In this view, a “practical, field-dependent, applied professional ethics” should be replaced by a “macroethics,” which is a theoretical, field independent, applicable ethics. That is, we conceptualize information ethics ontologically—information itself evolves from a simple category (“news or contents”) to a complex relationship, and, in doing so, changes the boundaries and ways in which we consider the ethical implications of it and its very nature. The “field” of information ethics is discursive. If we are to understand information ethics in Floridian terms, we are always at once discourse, in discourse. The challenge is getting outside “enough” in order to critique this discursive immersion. Information ethics is better understood as a process rather than a product.

If this is so—that is, if information ethics is process—we must reconsider how we teach it and research it. Now, this in many ways brings us back to the microethical level of the practical, the professional. If so, we need to be there, to understand where we are going. As information ethics scholars, we will need to reconsider content, pedagogy, and this process of IE itself. We will address the microethical issues but from within a heightened sphere of recognition, of awareness, and of being. This awareness will cross boundaries of all kinds. Take a “traditional” information ethics issue: privacy, and its “place” across the infosphere. There are different conceptualizations of this construct—cultural, generational, temporal, spatial, ideological, political, legal, professional, and so the list continues. These categories, of course, cut across each other, in a never ending mediation of interests. How they bear on specific information ethics issues is a complex debate, and it is often a great challenge to separate or sort them. And, quite possibly, some of the “traditional” issues of IE may simply be irrelevant in the emergent models of IE. Theoretically, privacy, for instance, must be considered within all of these to make any significant sense. And yet, to comprehend it, we must contain it with a narrow context. Is there something beyond “privacy” that makes more sense in this reconceptualization? Consider intellectual property and policy: it is clear that dominant paradigms of ownership and control lack the breadth and depth necessary to fit informational needs, as such cross borders, ideologies, politics, and ethics in general. Then, we again must reevaluate the problem and the process. There is, clearly, a tension between the micro- and macro-levels.

This tension currently reaches across the “discipline,” if we can call it that, of information ethics. We see the tension playing out as scholars who engage and collaborate, and perhaps clash, in this realm of discourse; scholars who previously were disconnected from one another by disciplinary boundaries are now entering the same spaces, coming from different micro-level perspectives, scholars united by their engagement, their mutual existences in the infosphere. What commonalities exist? Perhaps our mutual concerns here are all accommodated by the fluidity of the sphere. As simplistic as it sounds, we have much to learn from each other. We’ve reached a great point in information ethics—that we can move to a meta-discourse, where indeed we are engaging in a sophisticated discussion about our very being, our role in and as information ethics and ethicists. When Floridi proposes that we need a “whole new education and sensitivity to realize that the infosphere is a common space, which needs to be preserved to the advantage of all,” he is encouraging that we not only examine, but indeed reconstruct many long-standing boundaries. He is asserting change across a range of beliefs and structures. This change will impact a continuum of issues from pedagogy, disciplinarity, professionalism, culture, to being itself. The question looms, however, Can we institutionalize these changes? Should we? How would we? We are again back to the issue of responsibility. Indeed, we are charged with cultivating virtues of integrity and honesty, through consideration of the “other person” as a fellow human being, and balancing the rights of individuals, and the duties to society and to fellow professionals (Iacovina 2002; Buchanan and Henderson forthcoming). Still, we are facing newfound abstractions around the concept of the “other person,” through online personae, virtual worlds, avatars, bots, and so on. Our ethical obligations and responsibilities are reacting to them (see, for instance, Buchanan and Ess, on Internet research ethics).

Ultimately, this is no small charge. Continued discussions around social responsibility, professional responsibility, and personal responsibility must exist, and they will exist across those boundaries named above, if we allow ourselves to mediate the spaces afforded by the discourse of information ethics. Floridi’s is a call that resonates across boundaries, and a call very much worth heeding. It is our responsibility to continue the exploration.

References


Comment on “Understanding Information Ethics” by Luciano Floridi

Samir Chopra

affiliation?

Floridi suggests a new ethical stance via an “information ethics,” a “patient-oriented, ontocentric, ecological macroethics” that leads to a revised understanding of the traditional objects of our moral concerns. Floridi utilizes two concepts: the infosphere, “the...informational environment constituted by all informational entities...their properties, interactions, processes, and mutual relations” and re-ontologization: “a very radical form of re-engineering...that fundamentally transforms [a system’s] intrinsic nature.”

Amongst the taxonomy of fundamental trends in this re-ontologization is the “frictionless infosphere,” which results from barriers to information flows being broken down by a combination of technical and social protocols. But as the network neutrality problem shows, and as the digital divide ensures, legal, financial, and social factors continue to throw sand in the wheels of the infosphere. The results of a clinical trial on drugs are made available only to a select group of corporate pharmaceutical researchers; the code of
a scientific computation package is not made available to a group of researchers in a developing country; and legislation such as the DMCA ("thou shalt not write code that decrypts content-protection schemes") stifles the expressive content of software. Indeed, the very digitizing of information makes it more likely that technical locks can be placed upon it. Placing restrictions upon the traditional paper book is difficult; once I have purchased a copy of Crime and Punishment, the publisher is hard-pressed to restrict my reading habits or methods of sharing, but an e-book version of the same text is easily rendered unreadable and unshareable. The infosphere’s contours, then, continue to be shaped by a variety of forces, not all of which break down barriers to information flows.

Conversely, untrammeled information flows might hinder moral decision-making:

An explosion of the availability of information...sometimes leads to a socially irrational investment of resources...In the 1930s, welfare economists used the example of the factory that spews pollution onto the landscape, but pays nothing for it...If the full social costs of production were taken into account, the factory’s products would be uneconomical...maybe a future Pigou will write an analysis of the blindness of information economics at the close of the twentieth century and will point out that we were oblivious to the “information pollution” we were creating, that our economics did not force us to internalize the costs of overproduction, leaving us free to continue to pollute.4

As the economist Jacques Drèze2 and the psychologist Alex Bavelas noted in theoretical and experimental settings respectively, more information is not always a good.

Floridi suggests that information will become our ecosystem as increasingly more entities become networked and as we come to inhabit a[n]...environment of...persuasive, distributed...information processes.” Indeed, “the world itself...will be increasingly interpreted and understood informationally.” But these projections are only addressed to those that have access to the “digital cameras, laptops...mobile phones, camcorders, wireless networks...” that Floridi refers to. It is socio-economic circumstances that permit the ubiquity of this panoply of goods; the economy of material circumstances still affects the likelihood of the infosphere becoming our ecosystem. There still persist, then, prominent barriers to the realization of a full-blown informational environment.

But if we grant this, then our understanding of human moral agents as “interconnected, informational organisms” highlights their “intrinsically informational nature,” and since prior conceptualizations of information ethics are either too simplistic or too inclusive, information ethics is best understood as a macroethics that renders our ethical concerns ubiquitous and ambitiously universal: “being/information has an intrinsic worthiness...any informational entity has a Spinozian right to persist...ethical discourse concerns any entity, understood informationally...not only all persons...animals, plants...but also anything that exists...anything that may or will exist...anything that was but is no more.”

This, then, is a claim for an ethics whose scope of concern is far broader than we might have imagined, bringing in its wake a series of hard problems to be tackled. For these suggestions that our ethical concerns become ubiquitous might place a series of hard problems to be tackled. For these suggestions that our ethical concerns become ubiquitous might place a series of hard problems to be tackled. For these suggestions that our ethical concerns become ubiquitous might place a series of hard problems to be tackled. For these suggestions that our ethical concerns become ubiquitous might place a series of hard problems to be tackled. For these suggestions that our ethical concerns become ubiquitous might place a series of hard problems to be tackled. For these suggestions that our ethical concerns become ubiquitous might place a series of hard problems to be tackled.

Floridi suggests that in a fully realized infosphere, there will be no right to ignore, that our moral obligations will increase, for obligations only become effective when they are known. Are we then to be placed in a state of constant concern about moral failure of the kind that Michael Slote noted in his arguments for a virtue ethics to replace traditional moral theory?5 Slote’s concerns centered on the strain placed upon moral agents by traditional ethical theories by their seemingly excessive, self-sacrificing demands. In similar fashion, Henry Sidgwick had warned of the difficulties that inhered in moral codes that asked too much of their adherents; lest there be a general breakdown of compliance with morality.6

The ambit of our traditional ethical concerns is a small one, despite the exhortations from religious and secular humanist discourse that we broaden it. It requires considerable effort to get moral agents to look beyond their immediate concerns; how then are we to make the leap to the expansive notion at hand? This is not to suggest that this notion is absurd; there is ample precedent for it in ethical traditions such as those of Buddhism and the Advaita Vedanta, philosophies which sought to place us in a holistic context and in an ethic of care that was not chauvinistic. One welcomes this convergence of ancient moral traditions with contemporary ethical theorizing. Perhaps the most powerful device placed at our disposal by this analysis is that by an understanding of our informational natures, it provides a grounding for such expansive moral concern. We have, as it were, come to the logical conclusion of holistic ethical codes. But this current formulation is especially ambitious in its atemporal nature; we are to be concerned with the past, present, and future of all beings.

So, the path forward to the adoption of the ethical code at hand is not immediately clear. The precise manner in which such an information ethics may be formulated, and the guides to action that it could provide, present themselves as provocative challenges. What normative import does such an all-encompassing ethic have, and how much guidance might it provide us in daily ethical decision making? How are we to be guided in this new moral field, and how might we make the transition from our current ethical frameworks? If we are akin to the physicist, who is told that Newtonian mechanics is an incorrect theory and is to be supplanted by relativistic methods, but that for all practical purposes he may carry on as before for his work only concerns non-relativist domains, then have we been provided enough substantive and procedural guidance for our moral conduct?

The realization of the infosphere in the form envisaged by Floridi is a substantial challenge in itself, as is the problem of how best to utilize the potential cornucopia of informational wealth that is its inevitable result. I hope that my remarks above have also suggested that even more work remains to be done when we realize our informational natures and positions within the new ecosystem that we might come to inhabit.

Endnotes

Two arguments concerning intelligent, individualization of instruction are considered here. The first argument derives from an interview of Alfred Bork published in Educom Review (1999). In his interview, Bork maintained that education could be vastly improved on a global scale without high cost. This could be accomplished by means of adaptive, interactive, Socratic-style instruction delivered, not via the Internet, but via multimedia delivered on CD or DVD. My initial response was that adaptive, interactive, Socratic-style instruction requires artificial intelligence (which Bork denies) based upon empirical research, and these requirements make the cost prohibitive. Moreover, this increased cost cannot be recovered by developing an economy of scale, i.e., by recovering costs through serving more students. Why not? Well, the real challenge of individualization is to accommodate a wide range of individual differences. This range is increased when the total number of students is enlarged, particularly when this increase crosses geographical and cultural boundaries. The empirical research needed to identify pedagogically significant differences and develop effective means of addressing those differences greatly increases cost. Arguably, this increased cost involves a vicious circle: widening the range of students served in order to recoup costs produces a greater range of variation, which requires more funding for research and development, which must be recovered by increasing student demand, etc.

This argument concerning cost is important because justifications for technology development, particularly new technologies, are often based upon claims concerning quality, cost, and accessibility. If decreased cost is problematic, we are left with quality and accessibility and with balancing any gains with the costs in particular contexts. While increased accessibility has been heralded as one of the contributions of the Internet, my focus here will be upon gains in educational quality as afforded by empirical research and artificial intelligence. Historically, one main avenue of this research has been in the form of intelligent tutoring systems. Components of such systems often involve models of the student, the teacher, and the subject matter expert, and these models emulate expertise by means of production rules. Progress on adequate production rule systems has been exceedingly slow, and many have given up this endeavor. One alternative to production rule systems will be described below, and then another argument concerning instructional systems will be considered.

The attempt to develop useful hint mechanisms for students learning deductive proof construction has a long history. Production rules that model expertise in solving proof problems offer one approach to meeting this challenge. Given the difficulties of this approach, the CMU CTAT project attempts to automatically generate production rules from actual proof construction attempts. An alternative approach is to use records of actual student efforts to build a Markov Decision Process representation for use in generating helpful hints. Paths taken by students can be stored and rated in respect to their success or failure, and these in turn can be consulted when a student requests help in completing a proof. Hints can be given in varying degrees of helpfulness by conditionally specifying promising rules, premise expressions, and/or expression to be generated.

Both of these projects attempt to break the link between intelligent instruction and prohibitive cost. Yet, there remains a question and a related argument concerning the degree to which individualization is achievable. In theory, we can think of identifying and addressing the strengths and weaknesses of any individual student, but this idealization may not be achievable. Just how idiosyncratic individual needs are in particular educational contexts is yet undetermined. In any event, it may be best to design intelligent instruction on the assumption that, for any automated system of instruction, there is some student who is not well served by it. This assumption, and one that postulates human efficacy in identifying and addressing individual instructional needs, would lead to a different conclusion than that put forward by designers of automated instructional systems: namely, that humans should always be included as backups for any automated system of instruction. One corollary is that student models, databases, and related procedures for supplying individual feedback should be transparent to teacher, student, and whomever is involved in supplying the human backup.

I cannot vouch for the truth of the aforementioned assumptions, but as someone investigating the design of computer-based instruction, their acceptance on pragmatic grounds seems reasonable.

In sum, intelligent instructional systems should be designed with the expectation that not all students will be well served. Some direct, human backup system should monitor, identify, and address the needs of those students, and student models should be transparent and useable by human teachers (and perhaps students). The cost of building and maintaining such systems, or those proposed by Bork, should not be underestimated in any proposal to uplift education on a global scale. Designers often work towards a goal and accept its potential achievement as a working hypothesis. But having doubts about the ultimate success of the ideal aimed for can also move the design in a beneficial direction.

Endnotes

of computerized information allows it to reach many more students than traditional teaching can; however, there is much more to computer-aided education than just its size and speed. Take me, for instance: Would I be a philosophy major nearing graduation without an online program? Almost certainly not. But the reason has as much to do with the style and quality of online learning as it does with time and money. To give a brief account, my first education in the subject was at a private, multidisciplinary school in Santa Fe, New Mexico, called St. John’s College. If you’ve heard of it, you know that at St. John’s, classes are small, and discussion and writing are the heart of learning even technical subjects like language and mathematics. I did well at St. John’s and was very comfortable in the multidisciplinary environment, but I was unable to continue there after two years. I had every intention of continuing my education, but I didn’t count on what it would be like to take philosophy classes in a standard university classroom, where the normal class environment was fifty or more students, dominated by one-sided lectures rather than group discussions, and grades were determined far more by rote examination than any skill in argument, analysis, or paper-writing. I didn’t last a year in the university, and then, with a new baby and a full-time job to think of, I gave up for a while. It was a friend of mine who gave me back into the field with a Christmas gift; she bought me a single four-credit class at Harvard University’s online Extension School program. I found one I was interested in and took it, mostly for fun, but I didn’t expect it to be much like “real school.” I had figured, as many people do, that online classes were suitable only for subjects that could be mostly memorized, or learned without doing too much independent thinking, like locksmithing, veterinary medicine, or an MBA. We forget how flexible the Internet is, I think, and how much it favors communication; when people think of computer-enabled education, I think they envision sitting at HAL’s feet with a notebook, trying to keep up with a monotone recitation of facts.

Of the Harvard class, I expected a rigid environment of reading and regurgitating, like taking quizzes on the Internet, only with better subject-matter. I was very wrong. The class I took was comprehensive and challenging, with plenty of discussion and opportunity to learn from some very high-caliber professors and students. I learned so much in that one class that I immediately decided that I wanted to keep going this way and finish my degree, which lead me to the University of Illinois’ excellent online philosophy program. I still have a full-time job, and my daughter is four years old. Without the flexibility to do my studying at night, or on lunch-breaks, and to write my assignments and tests in windows of time (ranging from 24 to 72 hours), I would never be able to keep up. And while taking classes online is a more solitary experience than living in a bustling dorm full of like-minded students, it’s also some of the best education I’ve participated in, with challenging material, lots of feedback, and plenty of opportunity to specialize and expand on what’s learned in class.

The process of learning through an online interface is vastly different from classroom learning, and the need for some technical skill is only a superficial part of the difference. First of all, rather than happening at regular intervals a few times a week, it often takes place a little every day, at different times. This is convenient for busy people, but it can be daunting without good study and organizational skills. Also, the vast majority of communication between students, and between the student and the professor, takes the form of typed messages. This can be very positive, but it requires even more in the way of reading comprehension skills than normal study. One way in which the “forum” method that’s common to online education is useful is that it’s a format in which everyone has a place to develop his or her own views as well as a place to discuss and argue other students’ viewpoints. Also, unlike in a “meatspace” classroom, the professor can monitor and participate in every discussion, if he or she chooses (and has the stamina!). And there are even subtler particularities that are a direct result of the electronic method. For example, one cannot interrupt another student. It’s possible to review the professor’s comments several hours or days after they’re made. And things like it being a nice afternoon on a Friday don’t have the same impact on class as they do in real-time.

In understanding how these phenomena affect the educational process, it’s important to realize that they did not originate with advanced Internet technology—they’re part of the nature of communications devices.

II. A Rose Over the Telephone Would Not Smell As Sweet.

At the present time, computers still have a simple role in human education—they store information, and allow us to communicate increasingly complex ideas across greater distances than ever before. It would be a mistake, however, to think of computers as simply a pipeline, contributing nothing to the process of disseminating knowledge, applying thinking to it, and achieving understanding. No device that changes the nature of our communication is frictionless; always, the content undergoes some change because of the interface through which it is communicated. A good example is the telephone.

The telephone didn’t just enable people to have the same conversations they’d always had, only over larger distances; rather, fundamental aspects of conversation itself changed in response to the availability of the telephone. The telephone not only changed where we could be in relation to the people we were conversing with; it also, among other things, enabled conversations without the influence of facial expressions, body language, pheromones, and the potential for immediate physical contact. These changes had a deep impact on the way people interact, an impact that many of us are still not fully adjusted to. Many of us have witnessed an angry person yelling into a telephone, “I bet you wouldn’t talk like that if I came over there!” He is demanding to be treated as if there is the potential of physical contact in the conversation, even though in his situation, there isn’t.

Sometimes we don’t like to think that how we treat each other changes so much depending on the device(s) we’re using to communicate. But that’s only the tip of the cell-phone tower. Using computers as an interface for educating people has the potential to have a far more profound impact, not only on conversation, but on how we think and learn, and what it means to do both of those things.

III. Precedent Epistemological Changes: Reorganization of Information Storage.

Some of the major, foundational changes to what it means to “know something” are already visible as a result of the technology we have now. These are mostly shifts and expansions in what it means to “know”; one could call the overall effect an evolution. One such change is a result, specifically, of the end of a long era in which the human brain was the biggest efficient storage device available for our use. For most of our history, the ideal way to store knowledge was in the brains of smart people, who could memorize a lot of information and recall it quickly. The use of printed material expanded this capacity, but it was work-intensive to recover the data—you needed physical access to the books, not to mention the ability to read. So people with good “storage capacity,” as well as those with the ability to mine knowledge from books, were tapped
for jobs that required access to large amounts of raw data. This included being made into educators, with the job of spreading knowledge to other brains that would be capable (hopefully) of storing and using it efficiently.

Recently, however, the capacity of mechanical storage has expanded so much that it’s become more efficient to record information to magnetic drives or optical media than it is for smart people to memorize—it—more efficient, even, than our most modern methods of producing and disseminating non-electronic books. The fact that accessing a hard drive is still slower than accessing our minds (for however much longer this is a fact) is counteracted by the fact that mechanical storage can be scaled upwards indefinitely, and also that what is stored on a hard drive or optical device can be accessed by many minds, even simultaneously. If the drive is hooked to the Internet, then the information is, with no duplicative effort at all, instantly as accessible to others as it would have been if it had been published in a book that made its way into every library, or was read by professors in every school.

Currently, the paradigm of education still hinges on some old ideas about how much information a student should be expected to memorize, and what is the most efficient way to store and retrieve the data. Much will change as people learn to integrate the newer storage-and-retrieval mechanisms into the fabric of how they know things, and students integrate it more fully into their education. In order to do this, students will have to be taught how to use the Internet as a real source of knowledge, rather than be allowed to learn by trial-and-error or, as sometimes still happens, be discouraged from using it. In order to mine their new resources, they will need to learn, at a relatively early age, what we consider today to be excellent research skills. As more of the actual knowledge content that people need to learn becomes stored on external, computerized media, the more the role of educators will focus on teaching how to access, interpret, and understand that information, rather than on conveying it. (Obviously, a good educator both conveys and interprets information, but as storage and dissemination becomes more efficient, the access-interpretation-and-understanding aspect of teaching becomes more vital.) If teachers succeed in utilizing the power of computers to store and deliver information, the effect on education in the future could be hugely positive; if not, it will likely cause a disconnect between students and teachers that will make learning more difficult.

But regardless of how we manage the transition, what it means to know something will change. To “know something” used to include the fact of having access to that piece of information within our minds, as well as having some understanding of how it fits with other information. But now that our minds are being surpassed as the most efficient long-term method of storage, it’s beginning to make sense to expand how we think of “knowing” something to include learning a marker or a snippet of it; enough to induce the necessary level of understanding and lead one, when needed, to the correct stored content. As access to the stores of knowledge-content becomes more immediate (and hopefully more universal), with increasingly mobile technologies and eventually direct mental access to computers, less actual content will need to be stored internally. Instead, we’ll store mostly references and relationships and original thinking in the valuable space of our brains, and let the bulk of the raw data reside in manufactured storage devices.

In essence, technology will make room for us to know much more by “knowing,” in the sense of storing the information on-site in our bodies, less. The scope of “knowing” will expand—indeed, has already started to—to include “having the ability to access” the information. Understanding, however, which is a vital part of education, will continue to mean not only having access to the information, but being able to utilize it for practical and creative applications. And technology will have an impact on that aspect as well: computers can already help us determine when a person understands something, and their own synthetic powers are increasing. What it means for a human to understand something may not change, but it may not be for long that only the humans are capable of it.

IV. Precedent Epistemological Changes: Data Access and Education.

Related to the change in storage-mechanism is the quantity of information available at any time to average people. This quantity grows rapidly as technology advances, both “upward” in the sense of better technology, and “outward” in the sense of the technology being available to more people. Again, recently an important milestone has been reached: in just the past decade or two, it has become reasonable to assume that every first-world student has some type of access, often constant access, to the sprawling glut of information that is the Internet. It’s only a matter of time before access is instantaneous and hopefully universal. (Educationally speaking, it will be as important as schoolchildren today having access to books and the ability to read them.)

As this happens, the relative volume of what one person can “know” will increase dramatically, though how much depends on what kinds of freedom of access are maintained. Teachers will have to contend with, maybe in some ways compete with, a vast arena of information, probably conflicting information, that their students will all have access to. (Or, if not all of the students have the same access, this will raise its own problems.) Educators will have to focus less on presenting and accessing information—which was important in the textbook days of old—and more on being guides to finding, comparing, critically assessing, and understanding the huge amount of information which is already available.

V. The Relationship between Computer-Knowing and Human-Knowing.

So far we’ve seen how computer technology has caused some earthquakes, and will cause more, in the educational areas of information storage and information availability, both of which are fundamental ingredients in the process of human learning. But computers learn too, and their advancement has a definite impact on human knowing, just as human knowing has an impact on the progress of computer understanding. (In programming terms, this is known as a “feedback loop.”)

The paradigms of human learning and computer learning are not at all separate. Now we’re reaching the point where we and they have become smart enough to begin programming the computers to help program humans, or teaching them to teach us. And I don’t mean to simply equivocate between “programming” and “teaching.” Some types of modern “AI” software or Artificial Intelligence programs, though most experts do not consider them true, independent intelligences yet, do require an education that’s surprisingly similar to what a human receives. They start with a “neural net” or set of basic instructions, similar to the mental processes a child is born with, and require the input of informational content to further develop. They are designed to use the information they’re given, combined with their basic programming, to try and learn things they don’t already know. That’s not identical to how a human learns, but it is a process that humans use as part of their learning—which is, of course, why we decided to give it to computers as part of making them able to learn. On the other side of the coin, there’s no arguing that some aspects of
human education, especially at the early stages, bear a strong resemblance to the process of designing a computer program. Behavioral conditioning is a perfect example: our instincts (basic programming) combine with the information we’re given in a certain context to produce “programmed” responses. Again, this isn’t the whole picture, but it is indicative of the kind of cross-pollination that makes human and computer advances with respect to knowledge nearly inseparable.

To say it plainly, then, humans already do teach computers; and computers are already capable of one important part of teaching humans (providing access to the information). Computer-aided education will become more comprehensive and more commonplace eventually, and assuming that any increase in efficiency results from this practice (and it’s hard to see how it would not), then a cycle will develop: as computers advance, humans will get smarter through more and more available knowledge and probably improved delivery of that knowledge as well. The smarter humans will be able to program smarter computers, which will in turn help “program” smarter humans. This can be considered likely to happen no matter how radical or measured the changes to the human educational process are; whether it’s an overnight explosion or a gentle transitioning—in either case, we will change computers alongside ourselves, and in so doing provide a feedback loop that changes us further.

VI. Blurring Lines.

We know that computers will become more involved in human education, and that by doing so they cannot help but change how we understand and communicate what we know. But how much can something be tied to the process of our mental development, and extend our everyday mental ability, and still remain separate from us? At what point does the teacher-teaching-the-next-teacher relationship between educated humans and educating computers become a true symbiosis, a situation where the computer that teaches us how to think and stores the information we use to implement our thoughts—and maybe thinks of new ideas too—is more correctly thought of as part of our mind? At what point do we reach a level of entanglement between the computer-intelligence and human-intelligence galaxies where they’re no longer distinguishable for all intents and purposes?

To postulate that we’ll reach such a point is not as farfetched as it may seem—there are a lot of details, of different ways things could develop, between here and that point; and those details are mostly only mildly futuristic. For instance, does it matter if the system that our mind “incorporates” is a standard digital computer, an AI program, or a variety of biotechnological enhancement? The actual effects of computers on human knowing and thinking (and even identity) might be different in each case, but not qualitatively so. How different would humanity be if our ancestors had adopted bludgeons as their primary war-tools instead of swords? The answer is, “superficially.” Similar to the question of what the computers will actually be like is the question of how readily they’ll be adopted. As computers become more and more “part of us,” not everyone will embrace their new, expanded identity at first, and some never will—but if the technology is available, enough people doubtlessly will, and whole swathes of human behavior and biology will be altered as a result. Indeed, it doesn’t seem fantastical to postulate that we already are being altered.

Andy Clark, philosophy professor and author of Natural Born Cyborgs: Minds, Technologies and the Future of Human Intelligence, claims that a form of technological symbiosis began a long time ago, sometime in the prologue of our evolution perhaps, and that human beings have been adopting technology as an integral, biological part of our own thinking for centuries, at least. As ontological proof, he offers the fact that humans and some animals have an apparatus in our brains that give us a sense called “haptic touch,” whereby our brain adapts its “map” of our body to include a tool we are using. In this way, the shovel, conductor’s wand, or video-game controller in our hands becomes, as far as our neural network is concerned, an actual extension of our physical body. The human brain seems to be wired to not only adapt to its environment, but to actually incorporate it into the mind, effectively “plugging in” elements directly from the outside world into our mental milieu, just as if they were advanced prosthetic body parts, or microsofts that we plugged in behind our ears. Clark refers to this quality of human minds as “leakiness,” revealing that our identities are much “softer” and more inclusive of our environment than most people assume.

Following Clark, perhaps we already “identify” with our electronics far more closely than is usually acknowledged. This can be seen if we start by asking, “What makes a brain a part of our mental system?” One obvious answer is that if the brain is damaged, our thinking is impacted. Then it is worth asking, What happens to a person, especially one who’s been raised in the context of having constant access to external storage and processing power, if, say, their laptop blows up? Many young people would obviously find their ability to know and understand some things crippled, but even Andy Clark, a much older man, does not hesitate to compare the experience of losing his laptop to that of having a stroke, both of which he’s experienced. As he explains in an interview, he recovered from the stroke, “but the experience with the laptop was an awful lot worse at the time.” He lost more accumulated knowledge, and had to rebuild more of his intuitive structure for forming and organizing thoughts, when he lost his laptop than when he actually had a problem with his physical brain. Perhaps this is not normal, and maybe it never will be, but it certainly shows that computers can become an integral part of our mental systems.

The tendency of our minds to “leak” out into the environment and exploit it for what can honestly be described as computational shortcuts, seems to be evident in many cases beyond just computer technology. Consider scholars who can’t organize their thoughts without a pen and paper, musicians who come to need their instrument to express certain things that no other part of them can, and really almost any kind of craftsman you can think of in relation to his tools—the chef’s knife, the painter’s brush, the typists’ keyboard, the skateboarder’s deck. All of those things, besides being physical tools, can become augment to the thinking process, and once they do, losing them puts us at a disadvantage. In a sense, our minds are not just leaky—they’re lazy. They learn to rely on things in our environment that can take up some of the processing work-load of what we’re thinking and doing.

We aren’t potentially cyborgs in the distant future. We are potentially high-tech cyborgs, as the technology advances to that point, but we’ve been cyborgs for a very long time.

VII. The Little Leap Forward.

On this view, the advancement of computers into the realm of literally helping us think seems like a much smaller step. The idea that we might incorporate technology’s inherent ability to store information, produce and execute programs, and conduct rapid calculations into our own sphere of immediate knowledge (and maybe also conversely, that computers may inherit our ability to understand content and use it to think creatively)—that we might, in other words, co-habitate a thinking space with our technology—no longer sounds like science fiction, but only a matter of time and continued scientific effort.
And once computers become a bigger part of our education, it seems that their inclusion into our concept of what we need in order to think properly is unavoidable. When it’s the computer that gives you your assignments, that you turn to for research and ask questions for clarification, and that tells you how well your understanding is progressing and where you need to focus more effort, how can you not credit the computer for your education, just as the computer must credit humans for its education? The computer is taught by people, who eventually will have been taught by computers themselves—so, in effect, humanity will insert computer technology into its educational process as a sort of long-term springboard, functioning as an extra pulley, making the process less direct but exponentially more powerful. People (as a whole) will be smarter as the quality of computer system involved in their education increases because it will provide more and better access to knowledge content (as well as possibly many other helpful functions). They will use their increased intelligence to build smarter computers, which can then do a better job of teaching the next generation of human students, who will then be equipped to improve on the education of the computers that taught them, and so on. Throughout this evolution, humanity benefits from those tasks a computer can outperform us in, and derive the same benefit from us. As this process continues, the line between what’s us and what’s a tool we use to program ourselves, which is already blurrier than many people suspect, will certainly blur further. If the computer interface through which I was educated stays with me throughout my life, as future computers probably will; if the programming learns about and alongside me over time and is an ongoing, interactive component of my thinking, then how much does it really, functionally differ from my own brain? Is it not just a supplemental brain at that point?

Many tantalizing questions are still left—for instance, will the computers eventually develop personalities, making our intellectual co-existence with them something of a mental consortium, a sort of controlled, beneficial psychosis? Or will we, the “human side” of the new thinking organism, simply provide the personality aspect for the whole, perhaps resulting in personal computers that are an extension not only of their owners' operational knowledge, but of the social and interactive self as well?

We know how much of a leap in productivity and capability advanced technology has brought about for many areas of human enterprise: communication, research, manufacturing, entertainment, and so forth. Computer-aided education, though still a young enterprise with many kinks left to iron out, is obviously the way forward for our intellectual development.

VIII. Conclusion.

We already know that computer-aided education enables some people to pursue a good education who would not otherwise be able to. That education differs from classroom education in more ways than simply being long-distance, however, and many of those differences themselves are positive. The involvement of technology changes a great deal about the process, and will have further and more fundamental changes on the nature of education and the definition of human knowledge as time goes on: what it means to learn, as we understand it today, will contract, becoming more centered on having an understanding of something rather than having mental access to complete information about it; while what it means to know something will expand to include types of information-storage that are increasingly distant from our own brains. Changes in human knowledge will propagate themselves into the computers we build, an effect we can already see in the human-like “education” we give to some AI programs. As computers grow and become more important to education, human learning will advance in accordance with the advancement in the quality of computers we use to augment our thinking, knowing, and learning. This will spawn a cycle of increasing frequency, the overall effect of which will be to provide a springboard for the advancement of human intelligence.

Then, too, issues about the level to which human and computer identities will blur together, that seem merely academic and interesting now, will become increasingly relevant as changes and advancements propagate through the computer-human educational cycle. We already know that human beings exhibit a tendency to absorb tools into their biological concept of themselves for purposes of increasing mental efficiency, which seems to suggest that as computers become more integral components in our daily education and thinking activities, they will become more “part of us.” This means huge implications for how the definition of what it means to be human could change, implications that can seem equally frightening and exciting. Awareness and caution are in order, but fear of change is no reason to not take steps that can only increase our capabilities. We should take Andy Clark’s advice: “Know thyself, know thy technologies.”

Endnotes

1. or possibly “or”
2. This seems like it could have some scary consequences, but I think the common fears are minimally probable: Will we stop using our brains entirely, or grow hopelessly lazy? Unlike, since we didn’t grow “lazier” with the invention of the book. (In fact, that new method of storing and communicating more information made us noticeably more industrious.) Will we risk losing everything in some global power outage? Again, this is very unlikely; media can be copied and backed up almost infinitely, whereas it takes the (relatively easy) destruction of only a handful of human brains (or, say, the burning of a library at Alexandria) to wipe some knowledge from the Earth forever.

References


The Blended Classroom

Pat Manfredi
Southern Illinois University–Carbondale

Although some of us have taught or teach philosophy courses over the Internet, most of us teach exclusively on campus and will continue to do so for the foreseeable future. But the dichotomy between distance teaching and on campus teaching is a false one. Blended teaching incorporates many distance teaching techniques into on campus teaching without relying on web pages.

The digital syllabus. Regardless of whether a course is being delivered on campus or over the Internet, the course syllabus should be designed and used as a digital document.
Begin to think of your course syllabus as a digital document that can, if absolutely necessary, be printed, rather than a print document that is stored digitally. A student with a laptop has easier access to a digital syllabus than to a printed syllabus (which might be in the middle of a stack of papers somewhere in a dorm room). Students without laptops store course work on USB drives that can be used in campus labs or in their rooms. Thinking of a syllabus as something that students use from a computer encourages us to begin thinking of Internet resources that will be of use to students during the semester and that will supplement texts, lectures, and discussions. Providing hyperlinks to those resources in a syllabus gives students access to those resources directly from their word processor.

You probably already include links to your email address, your home page, and your department’s home page in your syllabus. But have you considered providing links to sites where texts are available for online purchases? More important are links to supplemental materials for a course: readings, interviews, podcasts, blogs, lectures, and conference proceedings. These can be listed in a syllabus under the day or week when their content is most relevant. To help students begin researching, provide links to your library’s catalog, interlibrary loan, and Google Scholar. Other resources, such as the Philosophers Index and JSTOR (a Scholarly Journal Archive) might be accessible only through links provided at your school library’s web site. Many students do not know about these resources and will not discover them on their own. Yes, students do need to learn to use the library. But, whether we like it or not, students typically check the Internet before going to the library. There is an enormous amount of trash that goes by the name “philosophy” on the Internet. But there are also many excellent resources and providing links to some of these helps students learn to distinguish quality academic web resources from worthless ones. The materials they find might keep a student from visiting the library, but they might also help to clarify and focus that student’s thinking so that time spent at the library is more productive. If nothing else, you might see less dependence on Wikipedia in your students’ papers.

**Distributing Course Materials.** Many professors use web pages to post lectures, handouts, and reading guides. Unfortunately, because the Web is primarily a graphical environment, too much text produces ugly, ineffective web pages. If your course materials are primarily texts and handouts, create word processing or PDF documents and send them to your students as email attachments. Because these documents can contain images and hyperlinks, you are not losing most of the advantages of the web. But you will save yourself considerable time since you will not need to format your documents in HTML. Plus, if you create an email group for the students in your course, you will only need to send one email. Alternatively, create a column for students’ email addresses in the spreadsheet that contains your class grades. This allows you to quickly select all students or a subset of students with whom you wish to communicate.

If you depend heavily on lectures and class discussions, use a voice recorder to record class meetings. This is particularly useful when students must miss class and would otherwise depend on another student’s class notes, or when students want to review material from a lecture. I am currently experimenting with an Olympus voice recorder (Model WS-210S) and an optional noise canceling lapel microphone. The recorder is somewhat larger than two packs of gum placed side by side. The recorder connects directly to a computer’s USB port and recordings (formatted as Windows Media Audio files) can be transferred to a computer like any other file, and distributed as email attachments.

If you want to push the technology, try setting up a podcast for your course. Podcasting does for audio programs what TiVo does for television programs. Individuals “subscribe” to podcasts and their content is automatically downloaded to their computers. Subscribers to a podcast receive programs as they are published, but listen to them at their leisure. Because podcasts are MP3 files subscribers can listen to the shows with their computers or transfer them to their iPods or other MP3 players. Podcasting offers the possibility for instructors to record lectures or seminars and to publish them, making them available for students taking the course on campus or from a distance. But podcasting content is not limited to audio files. Movies, images, documents, spreadsheets, and PowerPoint presentations can all be distributed by podcasting. A podcast course would be similar in many ways to a traditional correspondence course. Instead of cassette tapes and books being sent through third class mail, MP3 files and other documents would be sent over the Internet. For details, I recommend *Podcast Solutions: The Complete Guide to Podcasting* by Michael Geoghegan and Dan Klass.

**Collecting Assignments and Grading.** Try going paperless for your classroom assignments. I have not accepted printed assignments in any of my classes for over two years. I require all assignments to be created as Microsoft Word or RTF (rich text format) documents and to be sent to me as email attachments. This works surprisingly well and gives me the option of making the assignment deadline midnight on a particular day or before class begins (students love that!). Sophisticated paperless grading sites are available through Wadsworth and Turnitin, but I use the Track Changes and Reviewing features of Microsoft Word to mark the assignments as I read them. The track changes feature allows me to insert text with a color and format that I choose (say blue and underlined), and to mark deleted text with a color and format (say red and strikethrough). With Word’s reviewing tool bar I can insert comments about any selected portion of the text. The comments are numbered and recorded in a separate comment pane or in a “thought bubble.” Word highlights the text on which I have commented in light yellow and inserts a comment reference mark after it. When I return the assignments, students can either view the editorial changes and comments on the screen or print out the document with the changes and comments marked. Students without access to Microsoft Word can download Microsoft Word Reader for free to view my comments. Finally, if a student is required to revise a paper and return it to me, I can use the compare documents option to quickly compare the revised assignment with the original to see exactly which changes have been made.

Having students submit their papers through the Internet allows instructors to use Internet resources such as Turnitin (www.turnitin.com). Turnitin offers institutional and instructor licenses for tools that include paperless grading, an electronic grade book, peer review, digital portfolios, posting a class syllabus, and discussion boards. But it is most widely used for its plagiarism prevention tool. Instructors create a separate account within Turnitin for each class they teach and are then assigned a class ID and password. Using this ID and password, students log in to the site and submit assignments to be checked for originality against a database of documents that the company claims exceeds 4.5 billion pages. After a student submits a paper, Turnitin generates an originality report which matches text in the submitted document with text from documents available on the Internet and from other student papers that have been submitted. Matching text from a particular source is highlighted with a specific color and the percentage of text from the students assignment which matches that source is listed. If the source is available on the Internet, instructors can compare the source and a student’s essay side by side.

—— 26 ——
Class Discussions. Threaded discussion groups and blogs are a useful addition to many courses and they offer several distinct advantages: (1) they allow dialogue without an instructor’s presence; (2) multiple discussions can proceed simultaneously; (3) they permit discussions to continue throughout a semester; (4) they give a voice to students who are reluctant to speak in class; (5) they provide an environment where student comments can be more thoughtfully and more thoroughly developed; and, finally, (6) they encourage students to write for and to communicate with their peers rather than only with their professor.

But such groups are no substitute for face-to-face discussions and these cannot easily be reproduced through the Internet. Chat room environments share three features of classroom discussions: (1) they effectively engage students; (2) they allow multiple synchronous participation; and (3) they offer a quick response time. The problem is that most chat environments are text-based. This usually limits chat exchanges to phrases or short sentences and makes chatting with several individuals nearly impossible. Hosting an open chat session is nothing like moderating an in class discussion.

Like many instructors, when leading a classroom discussion I often act as a sort of filter and coach, rather than an advocate for some philosophical position. In my role as filter, I attempt to ensure that the discussion is not a free-for-all, but a learning experience that begins to model a good philosophical exchange. If a student asks a question, makes a comment, or replies to another student, I might help the speaker to clarify and to focus the remark and to eliminate parts of the remark that are irrelevant to the issue under discussion. Alternatively, I might help the class or the student to whom the comment is addressed correctly interpret the force of the remark and explore various replies.

But the complex sentence structures and quick response time in a conversation are very difficult to replicate in a text-based chat. Technology may eventually remedy these shortcomings, but, as things stand, the Internet cannot replicate the kind of learning that occurs during the give-and-take in a class discussion. Still, the Internet can help improve how we teach philosophy. Incorporating philosophical resources available on the Internet enriches the content of our courses. Requiring students to submit assignments electronically and embedding comments into those documents allows us to track improvements in students’ writing as well as increases in philosophical knowledge and sophistication. Using email to distribute outlines, handouts, and other types of course materials keeps students engaged in the course and frees class time for discussion—something that still is most effective when done face-to-face.

**Philosophy at The Open University**

**Derek Matravers**

**British Open University**

Philosophy seems to be flourishing in the United Kingdom at the moment. According to an article in The Education Guardian, the number of single honors students more than doubled between 2001 and 2006.¹

This is partly the result of the government policy of making more places available at university, and partly the result of an increase in popularity in the subject. The number taking the Philosophy ‘A-level’ (an examination available to those UK students who are still in school at eighteen) has recently been showing an annual increase of around 30 percent. This jump in popularity among the young is showing some signs of prompting an interest in those members of the adult population still interested in educating themselves. Fortunately, the UK is well-provided to cater for any such surge, should it materialize. The Open University (OU), which has had a philosophy department since its inception, is prominent in the adult education part-time sector.

The University was founded in 1969, in an attempt to broaden access to higher education in Britain. Students would be able to study part-time, from home, and were taught by means which included the cutting-edge of technology at the time: the television and the telephone. Since then, over two million people have studied with the OU, and over 625,000 qualifications have been awarded. Currently, there are over 200,000 students, with around 21,000 of those studying outside the UK (most of whom are in other European countries). There are, and this is unique for UK universities, no entry requirements whatsoever. The rationale of the institution is that people can have a second chance, no matter what their previous experience of education. The lesson of its success, which one can find either heartening or depressing, is that there are a lot more people who could benefit from a university education than have had the opportunity to do so.

The University is fully-integrated into the UK Higher Education sector. It swaps external examiners with other institutions (currently, members of the philosophy department are externals at Reading and Sheffield, and our externals come from Manchester and Cambridge). It is also judged in the way that any other UK university is judged. On many metrics, it scores very well. It has the highest rating for student satisfaction in both the 2005 and 2006 National Student Survey. Of the twenty-four subjects assessed by the Quality Assurance Agency (this measures teaching quality), seventeen were placed in the top “Excellent” category (philosophy scored a maximum 24/24). The Research Assessment Exercise is currently underway. Last time, in 2001, philosophy scored a 4 on a scale of 1-5* (5* being an improvement on 5).

The philosophy department runs programs at the undergraduate and graduate level. I will deal with them each in turn. Undergraduate students (of whom there are roundly eight hundred in any year) register for courses, rather than for programs. Courses (in the main) come in units of fifteen, thirty, or sixty “points” (it takes three hundred sixty points to obtain a BA(Hons)). If they studied continually and successfully, a student would take six years to complete his or her degree. Students are sent a box of material, which is a mixture of material written “in house” and publically available set texts. The standard courses of sixty points last thirty-two weeks and run from February to October. In each week, students are expected to study for eleven to fifteen hours. The backbone of the course is (usually) a course guide which provides pages of text, with in text exercises and self-assessment (increasingly done online). Students will also be referred to journal articles and other readings. At various fixed points over the year, students are required to submit an essay (usually there are six pieces of assessed work and an examination for each course). They also have a tutor who (ideally) lives locally, and will hold between six and twelve classes per year. The tutors are a mix of people: some coming to the end of their doctoral study; some retired academics, some people with good qualifications who have somehow never made it into a full-time academic post, and some who already have a full-time job, but enjoy the teaching. One of the sources of the enjoyment is, as anyone with experience in the sector knows, that there is no such thing as a “typical adult education student.” It is both exhilarating and alarming to think that one might be teaching justice to a high court judge, medical ethics to a surgeon, or (as once
happened to me) the philosophy of music to a conductor of a noted symphony orchestra.

The level of assessment strikes some as excessive. Perhaps because our founders were sensitive about how a university of this sort would be perceived, there is a tradition of heavy assessment and high standards. There is a ongoing debate as to whether we should drop one or more of the required essays, or award more first class marks (my view, for what it is worth, is that we could be more imaginative about the way we use our essays, and we could certainly be less mean in our marking).

Our founders have succeeded in their aims, however, as few if any doubt the quality of an OU degree. Looking at what happens to our students, many take advantage of the opportunities having a degree gives them in the workplace. (My favorite example of this is a reply from a student who I asked whether his degree would help him in his employment. He thought not, although he said that he would find it more relaxing given that he was no longer in danger of being fired if anyone found out he did not have a degree.) We “lose” several of our students to full-time study at other universities, including those (such as Cambridge) who have very stringent entrance requirements. Finally, to gesture at that specialist part of the market we know so well, even in my limited acquaintance I know of several full-time academics who began their academic careers as OU undergraduates.

There is a reasonably slow turn-over of courses; they are generally presented once a year for ten to twelve years. The courses are divided into first, second, and third level, which correspond to the standard expected at the three years of the usual British Higher Education Institution. Philosophy contributes to the interdisciplinary courses available at first level, and a new first level course on ethics will be presented next year. It has a single, broad-based course at second level, and currently two courses (one on political philosophy, one on the philosophy of mind) at third level. One of these has its final presentation next year and will not be replaced immediately. There is no single-honors degree in philosophy, although it is currently possible to take over half the points required for a degree in the subject, which qualifies students for a BA (Humanities and Philosophy).

The full undergraduate program is available to students in most EU countries. The OU offers some courses (although not philosophy) more widely; information on what is available where can be found at http://www3.open.ac.uk/contact/.

The philosophy department also offers a part-time MA in philosophy. This takes three years, although there are moves underway to move it to two years. This has been surprisingly successful: in any year, we usually have over two hundred MA students. It has the same availability as our BA courses. Successful MA students are welcome to apply for Ph.D. places at the Department, and, in addition, three or four each year move on to full-time graduate study at other universities.

Finally, the University offers both full-time and part-time research degrees. Currently, we have only eight graduate students. Departments no longer make use of external supervisors, which used to be a feature of OU graduate study. This meant that the OU keeps a closer relationship with the students, and spots problems earlier. However, it also puts a ceiling on numbers and has constrained the areas people can study to those in which some member of the Department is a specialist. Although the rules ask that students are “usually” resident in the UK, we are able to take good candidates who live abroad but travel to the UK at least twice a year. Indeed, the last two people we have accepted on the program live, respectively, in the USA and in Italy.

Doing a Ph.D. is always tough, and doing it part time, without the immediate context of a supportive academic community, is more so. Hence, completion rates do not compare favorably to institutions that do not teach part time and at a distance; only 62 percent complete within seven years (against a national benchmark of 74 percent). Furthermore, as our students are generally (although not exclusively) mature students with no particular interest in a career in the discipline, our record of placement is not good. There have been people with OU Ph.D.s who have got jobs, but not for some time. However, several of our successful Ph.D.s have had versions of their dissertations published. Two recent examples have been Gary Pendlebury’s *Action and Ethics in Aristotle and Hegel* (Ashgate, 2005) and Dave Edmonds’s *Caste Wars* (Routledge, 2006).

Currently, the OU is a little schizophrenic in its ambitions to be an international force in education. Some years ago, an attempt was made to set up a partner institution in the USA, which failed with the loss of a sizeable chunk of money. This has (understandably) made us rather risk-averse. However, we have partnerships with institutions around the world, whether that is the institution presenting OU material for its own degree courses, or the OU validating the quality of material they produce themselves. The OU was also awarded “Middle States Accreditation” (an endorsement of quality from the US), which has enhanced the status of OU degrees worldwide. Currently, the OU is thinking of looking abroad both for more partners, and also for making courses available globally that are taught (over the Internet) from the UK.

What of the future? The immediate horizon is not very bright. The government is making changes in the way it funds Higher Education that will have a damaging effect on OU finances. However, in other ways, the way distance education is going is interesting and exciting. The OU traditionally works by doing things on a large scale: good courses taught to hundreds of students. The model, as described above, is of a coterie of full-time academics who “produce” a course that is then “presented” by a local tutor. However, the distinction between production and presentation, long teetering, is now crumbling. The fact that the full-time academics are reachable by the students via email means that they can no longer shelter behind the local tutors. New technologies produce new ways of teaching: we can now produce lectures in the form of downloadable podcasts and also make use of blogs. This gives us an incredible flexibility; rather than taking three years to produce a course that does not change over its lifetime, blogs and podcasts can be changed in days or even hours. The days of the real-time lecture, accessible by thousands via their desktops, is not far away. The Department is already making use of this informally; one of our number, Nigel Warburton, produces podcasts of interviews with famous philosophers that has been a huge international success.²

There is every reason to think that the OU will continue to be a force in UK higher education, and some reason to think that it will become more of a force internationally.

Endnotes

1. http://education.guardian.co.uk/higher/story/0,2213665,00.html
BOOK REVIEW

Stepping Off the Edge of Reality


Tom P. Abeles
Editor, On the Horizon

“The task of philosophy is neither to explain the existing world (metaphysics) nor to change it..., but to construct alternative worlds, with different properties and foundations, to multiply the “world-ness” of existence... A philosopher thinks in worlds, and this may become a crucial capacity in the near future, with the proliferation of new world-like virtual realities and perhaps multi-dimensional physical realities... Through philosophical thinking a logical possibility of a new world will be posited and tested, and then...programmers will come to transform this possibility...” Mikhail Epstein, 1999

In the late 70s and early 80s the first virtual worlds, accessible on main frames, appeared. These text-based, non-graphic worlds for play and education gave rise to many of the games we see today and education and simulation platforms such as Diversity University. These were children of computer programmers, “techies,” and visionaries. Unlike Epstein’s vision, these “techies” and “gamers” were “products of the time” when flower-power and similar ideals scattered like petals by the spirits of the time. The rise of what one terms MMORPG’s or Massive Multiplayer Role Playing Games grew slowly until they exploded when the Internet opened up in 1995.

The unscripted worlds, or user designed environments, first gained high visibility as Second Life opened in 2003. These worlds are now developing as public spaces, controlled by the private sector creators/owners, or are being developed as special environments for a variety of uses, from educational platforms to simulations to try to understand a variety of social and environmental issues. But most of these are still seen as a variance of the familiar campus in brick space or a simulator of the spirits of the time. Few persons see these as represented by the movie The Matrix, a separate environment which has its own culture which is persistent and sufficiently stable to be considered the equivalent of “another country.”

Back in 1998, there appeared Michael Vlahos’s prescient article, “Entering the Infosphere,” wherein he suggests that the growing, networked world would create a virtual world wherein people could migrate in a fashion similar to the migrations across the planet. In other words, the reality would have sufficient substance that people could essentially commute between brick and click space, much like citizens commute through the crossings between the United States and Canada. About a year later, the movie The Matrix was released. Its plot basically had humans entrapped in a physical world but mentally transferred to a virtual world. Today, virtual worlds are one part of a thirty-five billion dollar global industry soon to dwarf in revenue the motion picture entertainment complex.

In academia, there has been considerable interest in a variety of disciplines at the individual level such as with virtual humans and embodiment of humans in technologies. Few, as Epstein suggests, have explored this area as “worlds” or systems. Stephenson’s “Snow Crash,” a science fictional development of a “metaverse” has become the paradigmatic example which has captured the attention of academics from a variety of disciplines, from hardware/software developers to social scientists and humanities scholars. But, these worlds largely have moved out of the realm of fiction into the arena of the technologists and entrepreneurs with the creation of games/simulations and role playing multiplayer environments. Philosophers, as Epstein has hoped for, have not seen this as a legitimate domain for scholarly study. As Epstein suggests, there is still a tailing effect within the discipline and the early adopters have not materialized in traditional philosophical domains.

This, of course, has been a significant problem within academia which operates within narrow domains and is seemingly unable to step outside of accepted boundaries. Edward Castronova’s paradigmatic contribution as an academic was to publish his studies of these virtual worlds as an economist. But these appeared largely outside of mainstream research. Synthetic Worlds, his first book, is a classic compilation of the economic underpinnings of these virtual worlds and a contribution which finally legitimized this as an arena for scholarly research.

In these synthetic worlds a vibrant economic ecosystem exists and is linked to the brick space world by a market for these virtual objects and currencies, which can be bought and sold for U.S. dollars and other brick space currencies. Thus, Synthetic Worlds set the stage for looking at these worlds not as games and entertainment, but as environments, much as projected by Vlahos and others, as new territories, essentially extending the area of planet Earth and opening up new territories for exploration and development.

In Synthetic Worlds, Castronova suggests that the designs of these worlds should be the responsibility of the social scientist, thus claiming what both Epstein and Vlahos believe might rightly fall to the arena of philosophy, reminiscent of Plato’s Republic, or the writings of Karl Marx. In fact, though, as Castronova points out, the development has already defaulted to the community of entrepreneurs.

In the end, Castronova goes astray with Exodus to the Virtual World. Rather than sticking with his critical analysis and insights about these worlds, he has attempted to seize the bully pulpit as a public intellectual. Rather than seeing these as two territories which influence each other as happens today, on planet, internationally, he attempts to politically and socially terraform brick space by importing all that he sees as positive in the Edenic virtual sphere.

Castronova clearly points out that the current virtual environments are being designed by game developers focused on a revenue stream which is dependent on participants paying fees to play in these environments. Thus, he painstakingly describes the various development domains within these corporations that are committed to creating entertainment, or, as Castronova says, fun. One needs to have fun and, thus, the experience must be pleasurable. That requires care in constructing and maintaining the worlds. Castronova creates the metaphor of “ministries” such as a justice and culture, or legal systems to maintain order, and the equivalent of a treasury department which concerns itself with the economic models under which the worlds operate.

But one needs to remember that unless the participants find the world interesting, they vote by taking their credit cards and going to new territories. Thus, what we have is a benevolent dictatorship or corporate run worlds which compete for attention in the form of monthly payments much like cable and satellite companies compete for eyes on the channels that they provide to viewers.
And here is where Castronova goes off track, throwing away his keen analytical insights as an academic and economist and putting on the cap of an Orwellian social planner. He has completely ignored the fact that a ten billion and exponentially growing industry is not only attractive to the venture capital community but to the corporate world that sees these worlds as markets, “in world” and as a medium for promoting their current brick-space products/services. Already, one virtual world developer has combined with a major network crime show, CSI: NY, to offer audience participation in both click and brick space. Second Slice is a virtual publication aimed at helping businesses market more effectively in Second Life and elsewhere.

The benevolent creators of the virtual worlds are sensitive to the idea that these experiences are fun, not because the developers are altruistic gods in their Olympian heights, but because they are, like operators of amusement parks, creating experiences that keep the customers returning whether for cotton candy or a heart-in-throat experience of dropping vertically from a tower at free-fall speeds. They are renaissance fairs whether one is dressed as a knight or “Spaceman Biff.” Do those who live on the planet want or will they accept the benevolent masters of the virtual worlds as masters of the planet? While one can move between worlds with a credit card, one can not, as of this moment, leave the planet permanently and escape to a choice of planets, real or virtual. The answer is “maybe.”

Exercises have been run with groups from young adults to senior citizens. Basically they have been asked to come to a consensus on five to ten ideas which would help make this planet a perfect world. When asking how one could actualize these ideas, only one answer has been forthcoming, the intervention of a power that would clean up our current mess and set us back on Earth. In other words, the solution in brick space is to create a controlling force that manages the world—the exact situation as with virtual worlds, a benevolent dictatorship. There is no hope in these groups that humans as a society will be able to restore us to some idyllic Eden where life is “fun.” Even in the movie The Matrix, the character Cypher chooses the taste of steak in his virtual world, though he knows that it’s not real because he knows that the physical world, on planet, is not “fun.”

Castronova is on target in his analysis of fun. And, he, like Vlahos and others, understands clearly that virtual worlds and brick space influence changes in each of the spheres. The idea of developing a hedonist theory as a basis for understanding how and why people migrate to virtual worlds and how this might affect changes on the planet are interesting but lack the serious analysis that Castronova can offer when he wears his economist’s hat. The planet is a work in progress. It’s not the blank canvas used by corporate developers of virtual worlds. There are players beyond human control on the planet, as we know from the issues surrounding global warming. Resources are not evenly distributed or even under uniform control, and the value systems of the world’s “players” are not those of the spectrum of folk playing in virtual worlds.

Castronova, in his rush to claim the bully pulpit to spread his message of “fun” as a way to rescue the social/economic and biophysical elements of the planet, clearly points to why John Brockman, in his Edge blogs and publications, has claimed the stage for the science community. The humanities in general, and philosophers in particular, have defaulted and given up the equivalent of their birthrights as setting the moral agenda for human kind. The Enlightenment failed to find a way to adapt the scientific method as a means for studying and managing the social sphere. And economists, as the current fight between the neoclassical and heterodox has shown, have not yet found the magic key either. The body of literature, from Vonnegut’s Harrison Bergeron to Orwell’s 1984, warns us about benevolent hands controlling the environment. It may work in cyber space, but even here, Castronova is far from convincing.

When people migrate to new territories they take their culture with them and it changes when it encounters a new environment. But the reverse has not happened to the extent that Castronova would wish to see happen. In fact, for most people, virtual space has been a place to experiment in order to find alternatives on the planet. Simulations of business, war, or social systems are examples. It has been difficult, except in the large MMORPG’s, to find self-sustaining cultures, much closer to the idea of a separate world.

Second Life and other user-created worlds are attempts, in many ways, to escape the limitations of brick space. It is this separate world that settlers in the New World sought. There was never an idea that this was an experimental world to show the “Old World” how to transform itself. Unfortunately, Castronova, after carefully analyzing these new worlds of “fun” cannot help himself to turn proselytizer for changing planet Earth without clearly addressing the many issues that are not addressed by the simplified corporate controlled virtual spaces, the virtual amusement parks where the illusion of “fun” is sold. Castronova has strayed from his strengths in this volume.

Once one has been led to the podium because of one’s academic expertise, there is a proclivity to speak almost ex cathedra on subjects outside of one’s discipline. The literature on the future of our planet is large and growing with the rise of concern over climate change. Philosophers to science fiction writers have researched and written about the possibilities of a return to an Eden or similar community. Exodus to Virtual Communities will not reside next to Plato’s Republic, Marx’s Das Kapital, or Stephenson’s Snow Crash.

Endnotes

NOTE

The 2008 North American Computing and Philosophy Conference, July 10th-12th at Indiana University

Anthony F. Beavers
The University of Evansville

The International Association for Computing and Philosophy (http://ia-cap.org) is sponsoring its annual North American conference this July 10-12 at Indiana University in Bloomington, Indiana. As with last year’s conference, held at Loyola University, this one will feature a mix of presentations on formal topics in computer science and cognitive science and on ethical matters pertinent to the conference theme. Last year, the IACAP examined issues surrounding the Free and Open Source Software (FOSS) movement and the Open Access (OA) movement in academic publication. This year, the IACAP will look at an equally provocative theme, the limits of computation, taken broadly to include not only issues surrounding formal
computation, but also those that examine what we can and cannot do with computers.

On the more formalistic side, the Herbert A. Simon Keynote Address will be delivered by the celebrated philosopher and cognitive scientist Paul Thagard (University of Waterloo), who will address the question of whether computers can understand causality, “using a neurocomputational account of how sensorimotor schemas underline human causal thinking.” Thagard’s past research on cognition ranges from inference and induction in 1986 to his most recent work on emotion.

Thagard’s keynote will be followed by a special session partially funded by the National Science Foundation’s CreativeIT program, which “makes possible an exploratory investigation of human creativity in the area of computer programming, with the hope of exploiting study of human creativity in order to eventually make significant contributions to automatic programming,” according to session co-chair and Director of the North American division of the IACAP, Selmer Bringsjord (Rensselaer Polytechnic). Bringsjord and his session collaborator, Konstantin Arkoudas, hope to transform the session, titled “Automatic Programming and Human Creativity,” into a book or special journal issue on the same theme.

The Douglas C. Engelbart Keynote Address, named in honor of the famed creator of the computer mouse, will be delivered this year by Ronald Arkin (Georgia Tech) and will concern “Ethics and Lethality in Autonomous Robots.” Arkin’s research over the course of his career has dealt with many dimensions of robotics. Most recently, he has been treating the topic of robot ethics. In his talk, Arkin will deal with concrete applications.

Robot architectural design recommendations will be presented for (1) post facto suppression of unethical behavior, (2) behavioral design that incorporates ethical constraints from the onset, (3) the use of affective functions as an adaptive component in the event of unethical action, and (4) a mechanism in support of identifying and advising operators regarding their ultimate responsibility for the deployment of such a system.

Arkin’s research is funded by the Army Research Office.

Following Arkin’s talk on ethics and autonomous robots, the conference will feature a special two-hour panel session asking whether ethics can be computed. Organized by Michael Anderson (University of Hartford), this panel will be populated by people who are well versed in the issues, some of whom are largely responsible for introducing philosophy (at least) to the moral problems and perils, not only of autonomous robots, but also of computers more generally. In addition to Michael Anderson and Ronald Arkin himself, the panel members will include Colin Allen (Indiana University), Susan Anderson (University of Connecticut), Marcello Guarini (University of Windsor, Canada), Tom Powers (University of Delaware), John Sullins (Sonoma State University), and Jim Moor (Dartmouth College), who also contributed greatly to the early organization of the system of conferences now sponsored by the IACAP.

The IACAP is grateful to Indiana University for hosting this year’s NA-CAP conference. The setting in Bloomington is particularly appropriate, given the contributions to computing and philosophy that have emerged from this institution over the years by such figures as Douglas Hofstadter, John Barwise, Andy Clark, and Timothy van Gelder. Bloomington remains home to Hofstadter’s Center for Research on Concepts and Cognition, and Indiana University’s Cognitive Science Program and its School of Informatics continue to provide a rich climate of exploration concerning the philosophical issues raised by modern computing practices. The conference organizers are planning a special opening session to highlight recent work on campus that ties to the conference theme.

The submission deadline for the conference is March 1st. So, it is difficult to summarize the rest of the program at this time, although conference organizers are hoping for paper submissions on what we can and cannot do with computers in areas outside of causal reasoning, automatic programming, and ethical decision making as well. Topics for other sessions may range broadly from addressing whether the computational metaphor is helpful or harmful for describing cognition to whether or not a musical or literary composition written by a computer can be considered as a work of art. The impact of computing on politics and social networking is also a possible theme.