OPEN PROBLEMS IN THE PHILOSOPHY OF INFORMATION

LUCIANO FLORIDI

Abstract: The philosophy of information (PI) is a new area of research with its own field of investigation and methodology. This article, based on the Herbert A. Simon Lecture of Computing and Philosophy I gave at Carnegie Mellon University in 2001, analyses the eighteen principal open problems in PI. Section 1 introduces the analysis by outlining Herbert Simon's approach to PI. Section 2 discusses some methodological considerations about what counts as a good philosophical problem. The discussion centers on Hilbert's famous analysis of the central problems in mathematics. The rest of the article is devoted to the eighteen problems. These are organized into five sections: problems in the analysis of the concept of information, in semantics, in the study of intelligence, in the relation between information and nature, and in the investigation of values.

Keywords: artificial intelligence, computer ethics, David Hilbert, information, knowledge, philosophy of information, semantics, Herbert Simon, information theory.

1. Herbert Simon's View

In October 2000, Carnegie Mellon University named the new computer science building the Newell-Simon Hall. On that occasion, a journalist interviewed Herbert Simon about the ways in which computers will continue to shape the world. Simon stated that "technology expands our ways of thinking about things, expands our ways of doing things." He then added, "Knowing a lot about the world and how it works. That's a major place where computers come in. They can help us think" (Spice 2000). These remarks are indicative of Simon's broad interest in the theoretical and applied issues emerging from the philosophy of computing and information (see Floridi 2004 for a review of the field). Simon was right in both cases. In 1962, he had already envisaged the future role of computers as conceptual laboratories, a valuable approach now widespread among researchers in the field (Simon 1962; Grim, Mar, and St. Denis 1998). On the other hand, this article could be read as a comment on Simon's first remark.

Technology unveils, transforms, and controls the world, often designing and creating new realities in the process. It tends to prompt original ideas, to shape new concepts, and to cause unprecedented problems. It usually embeds but also challenges ethical values and perspectives. In short, technology can be a very powerful force for intellectual innovation, exercising a profound influence on how we conceptualize, interpret, and transform the world. Add to that the fact that the more ontologically powerful and pervasive a technology is, the more profound and lasting its intellectual influence is going to be. Recall that technology has had an escalating importance in human affairs at least since the invention of printing and the scientific revolution. It becomes obvious why the conceptual interactions between philosophy and technology have constantly grown in scope and magnitude, at least since Galileo's use of the telescope.

The modern alliance between *sophia* and *techne* has reached a new level of synergy with the digital revolution. Since Alan Turing's seminal work, computational and information-theoretic research in philosophy has become increasingly fertile and pervasive, giving rise to a wealth of interesting and important results (see Mitcham and Huning 1986; Bynum and Moor 1998 and 2003; Colburn 2000; Floridi 1999, 2003, 2004c, and 2004 for references). Indeed, in 1998, introducing *The Digital Phoenix: How Computers Are Changing Philosophy*, Terrell Ward Bynum and James H. Moor acknowledged the emergence of a new force in the philosophical scenario:

> From time to time, major movements occur in philosophy. These movements begin with a few simple, but very fertile, ideas—ideas that provide philosophers with a new prism through which to view philosophical issues. Gradually, philosophical methods and problems are refined and understood in terms of these new notions. As novel and interesting philosophical results are obtained, the movement grows into an intellectual wave that travels throughout the discipline. A new philosophical paradigm emerges. [. . .] Computing provides philosophy with such a set of simple, but incredibly fertile notions—new and evolving *subject matters*, *methods*, and *models* for philosophical inquiry. Computing brings new opportunities and challenges to traditional philosophical activities [. . .] computing is changing the way philosophers understand foundational concepts in philosophy, such as mind, consciousness, experience, reasoning, knowledge, truth, ethics and creativity. This trend in philosophical inquiry that incorporates computing in terms of a subject matter, a method, or a model has been gaining momentum steadily. [1998, 1]

In Floridi 2003 I define this area of research as the *philosophy of information* (PI).

PI is a new philosophical discipline, concerned with (a) the critical investigation of the conceptual nature and basic principles of information, including its dynamics (especially computation and information flow), utilization, and sciences, and with (b) the elaboration of information-theoretic and computational methodologies and their application to philosophical problems.

A genuine new discipline in philosophy is easily identifiable, for it must be able to appropriate an explicit, clear, and precise interpretation of the classic "ti esti" question, thus presenting itself as a specific "philosophy of." "What is information?" achieves precisely this. However, as with any other field question (consider for example "What is knowledge?"), "What is information?" only demarcates a wide area of research; it does not map out its specific problems in detail. And a new discipline without specific problems to address is like a car in neutral: it might have enormous potentialities, but there is no progress without friction.[1] So the question that needs to be addressed is this: What are the principal problems in PI that will deserve our attention in the coming years? Or, to paraphrase Simon's words, how will ICT (information and communication technologies) expand our philosophical ways of thinking?

Trying to review future problems for a newborn discipline means looking for possible difficulties. Complete failure is one. Poor evidence, lack of insight, inadequate grasp of the philosophical situation, human fallibility, and many other unpredictable obstacles of all sorts can make a specific review as useful as a corrupted file for an old-fashioned program. Another trouble is partial failure. The basic idea might be good, the direction even correct, and yet the choice of problems could still turn out to be embarrassingly wide of the mark, with egregious nonstarters appointed to top positions and vital issues not even short-listed. And as if all this were not enough, partial failure may already be sufficient to undermine confidence in the whole program of research, thus compromising its future development. After all, philosophy is a conservative discipline, with controversial standards but the highest expectations, especially of newcomers. Added to this, there is the Planck Effect (Harris 1998). Max Planck once remarked:

> An important scientific innovation rarely makes its way by gradually winning over and converting its opponents: it rarely happens that Saul becomes Paul. What does happen is that its opponents gradually die out, and that the growing generation is familiarized with the ideas from the beginning: another instance of the fact that the future lies with youth. [1950, 97]

If the Plank Effect can be common in physics, imagine how it might be in philosophy.

Given the risks, is this visionary exercise really a game worth the candle? Arguably, it is. A reliable review of interesting problems need be neither definitive nor exhaustive. It does not have to be addressed to all our colleagues and can attract their graduate students. And it fulfills a necessary role in the development of the field, by reinforcing the identity

[1] "As long as a branch of science offers an abundance of problems, so long is it alive; a lack of problems foreshadows extinction or the cessation of independent development. [. . .] It is by the solution of problems that the investigator tests the temper of his steel; he finds new methods and new outlooks, and gains a wider and freer horizon" (Hilbert, 1900).

of a scientific community (the Wittgenstein Effect),[2] while boosting enthusiasm for the new approach. Obviously, all this does not mean that we should not go on tiptoe in this minefield. Looking for some guidance is also good idea. And since nobody has performed better than Hilbert in predicting what were going to be the key problems in a field, I suggest we first turn to him for a last piece of advice before embarking on our enterprise.

2. David Hilbert's View

In 1900, Hilbert delivered his famous and influential lecture in which he reviewed twenty-three open mathematical problems "drawn from various branches of mathematics, from the discussion of which an advancement of science may be expected" (this quotation and all subsequent Hilbert quotations are from his 1900). He introduced his review through a series of methodological remarks. Many of them can be adapted to the analysis of philosophical problems.

Hilbert thought that mathematical research has a historical nature and that mathematical problems often have their initial roots in historical circumstances, in the "ever-recurring interplay between thought and experience." Philosophical problems are no exception. Like mathematical problems, they are not contingent but timely. In Bynum and Moor's felicitous metaphor, philosophy is indeed like a phoenix: it can flourish only by constantly reengineering itself and hence its own questions. A philosophy that is not timely but timeless is likely to be a stagnant philosophy, unable to contribute to, keep track of, and interact with cultural evolution, and hence to grow.

Good problems are the driving force of any intellectual pursuit. Now, for Hilbert, a good problem is a problem rich in consequences, clearly defined, easy to understand, and difficult to solve, but still accessible. Again, it is worth learning the lesson, with a further qualification: genuine philosophical problems should also be *open*, that is, they should allow for genuine and reasonable difference of opinion. Throughout its history, philosophy has progressively identified classes of empirical and logico-mathematical problems and outsourced their investigation to new disciplines. It has then returned to these disciplines and their findings for controls, clarifications, constraints, methods, tools, and insights. Philosophy itself, however, consists of conceptual investigations whose essential nature is neither empirical nor logico-mathematical. In philosophy, one neither tests nor calculates. On the contrary, philosophy is the art of designing, proposing, and evaluating explanatory models. Its critical and

[2] "This book will perhaps only be understood by those who have themselves already thought the thoughts which are expressed in it—or similar thoughts. It is therefore not a text-book. Its object would be attained if it afforded pleasure to one who read it with understanding." Wittgenstein, *Tractatus Logico-Philosophicus*, opening sentence.

creative investigations identify, formulate, evaluate, clarify, interpret, and explain problems that are intrinsically capable of different and possibly irreconcilable solutions, problems that are genuinely open to reasonable debate and honest disagreement, even in principle. These investigations are often entwined with empirical and logico-mathematical issues and so are scientifically constrained; but in themselves they are neither. They constitute a space of inquiry broadly definable as normative. It is an open space: anyone can step into it, no matter what the starting point is, and disagreement is always possible. It is also a dynamic space, for when its cultural environment changes, philosophy follows suit and evolves.

Open problems call for explicit solutions, which facilitate a critical approach and hence empower the interlocutor. In philosophy we cannot ask

> that it shall be possible to establish the correctness of the solution by means of a finite number of steps based upon a finite number of hypotheses which are implied in the statement of the problem and which must always be exactly formulated

but we must nonetheless insist on clarity, lucidity, explicit reasoning, and rigor:

> Indeed the requirement of rigour, which has become proverbial in mathematics, corresponds to a universal philosophical necessity of our understanding; and, on the other hand, only by satisfying this requirement do the thought content and the suggestiveness of the problem attain their full effect. A new problem, especially when it comes from the world of outer experience, is like a young twig, which thrives and bears fruit only when it is grafted carefully and in accordance with strict horticultural rules upon the old stem.

The more explicit and rigorous a solution is, the more easily can it be criticized. Logic is only apparently brusque. The real trap is the false friendliness of sloppy thinking.

At this point, we should follow Hilbert's advice about the difficulties that philosophical problems may offer, and the means of surmounting them. First, if we do not succeed in solving a problem, the reason may consist in our failure to recognize its complexity. The accessibility of a problem is a function of its size. Philosophy, like cooking, is a matter not of attempting all at once but of careful and gradual preparation. Even the most astonishing results are always a matter of thoughtful choice and precise doses of the conceptual ingredients involved, of gradual, orderly, and timely preparation and exact mixture. The Cartesian method of breaking problems into smaller components remains one of the safest approaches. Second, it is important to remember that negative solutions, that is, "showing the impossibility of the solution under the given hypotheses, or in the sense contemplated," are as satisfactory and useful as positive solutions. They help to clear the ground of pointless debates.

So far Hilbert. A word now on the kind of problems that are addressed in the following review. To concentrate our attention, I have resolved to leave out most metatheoretical problems, like "What is the foundation of PI?" or "What is the methodology fostered by PI?" This is not because they are uninteresting but because they are open problems *about* PI rather than *in* PI and deserve a specific analysis of their own (Floridi 2003). The only exception is the eighteenth problem, which concerns the foundation of computer ethics.

I have also tried to focus on philosophical problems that have an explicit and distinctive informational nature or that can be informationally normalized without any conceptual loss, instead of problems that might benefit from a translation into an informational language. In general, we can rely on informational concepts whenever a complete understanding of some series of events is unavailable or unnecessary for providing an explanation (this point is well analyzed in Barwise and Seligman 1997). In philosophy, this means that virtually any question and answer of some substantial interest can be rephrased in terms of informational and computational ideas. This metaphorical approach, however, may be deleterious, for it can easily lead to an information-theoretic equivocation: thinking that since x can be described in (more or less metaphorically) informational terms, then the nature of x is genuinely informational. The equivocation makes PI lose its specific identity as a philosophical field with its own subject. A key that opens every lock only shows that there is something wrong with the locks. Although problems can acquire a new and interesting value through an informational analysis, the main task of PI is to clarify whether a problem or an explanation can be legitimately and fully reduced to an informational problem or explanation. In PI, informational analysis provides a literal foundation, not just a metaphorical superstructure. The criterion for testing the soundness of the informational analysis of a problem p is not to check whether p can be formulated in informational terms but to ask what it would be like for p not to be an informational problem at all.

With the previous criterion in mind, I have provided a review of what seem to me some of the most fundamental and interesting open questions. For reasons of space, even those selected are only briefly introduced and not represented with adequate depth, sophistication, and significance. These macroproblems are the hardest to tackle, but they are also the ones that have the greatest influence on clusters of microproblems to which they can be related as theorems to lemmas. I have listed some microproblems whenever they seemed interesting enough to deserve being mentioned, but especially in this case the list is far from exhaustive. Some problems are new, others are developments of old problems, and in some cases we have already begun to address them, but I have avoided listing old problems that have already received their due philosophical attention. I have not tried to keep a uniform level of scope. Some problems are very general, others more specific. All of them have been chosen because they indicate well how vital and useful the new paradigm is in a variety of philosophical areas.

I have organized the problems into five groups. The analysis of information and its dynamics is central to any research to be done in the field, so the review starts from there. After that, problems are listed under four headings: semantics, intelligence, nature, and values. This is not a taxonomy of families, let alone of classes. I see them more as four points of our compass. They can help us to get some orientation and make explicit connections. I would not mind reconsidering which problem belongs to which area. After all, the innovative character of PI may force us to change more than a few details in our philosophical map. And now, to work.

3. Analysis

> The word *information* has been given different meanings by various writers in the general field of information theory. It is likely that at least a number of these will prove sufficiently useful in certain applications to deserve further study and permanent recognition. *It is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of this general field.* [From "The Lattice Theory of Information," in Shannon 1993, 180, emphasis added]

Let us start by taking the bull by the horns:

P.1: The elementary problem: What is information?

This is the hardest and most central question in PI. Information is still an elusive concept. This is a scandal not in itself but because so much basic theoretical work relies on a clear analysis and explanation of information and of its cognate concepts. We know that information ought to be quantifiable (at least in terms of partial ordering), additive, storable, and transmittable. But apart from this, we still do not seem to have a much clearer idea about its specific nature.

Information can be viewed from three perspectives: information as reality (for example, as patterns of physical signals, which are neither true nor false), also known as *ecological* information; information about reality (semantic information, alethically qualifiable); and information for reality (instruction, like genetic information). Six extensionalist approaches to the definition of information as reality or about reality provide different starting points for answering P.1:

- the communication theory approach (mathematical theory of codification and communication of data/signals, Shannon 1948;

Shannon and Weaver 1949) defines information in terms of probability space distribution;

- the probabilistic approach (Bar-Hillel and Carnap 1953; Bar-Hillel 1964; Dretske 1981) defines semantic information in terms of probability space and the inverse relation between information in *p* and probability of *p*;
- the modal approach defines information in terms of modal space and in/consistency (the information conveyed by *p* is the set of possible worlds excluded by *p*);
- the systemic approach (situation logic, Barwise and Perry 1983; Israel and Perry 1990; Devlin 1991) defines information in terms of states space and consistency (information tracks possible transitions in the states space of a system);
- the inferential approach defines information in terms of inferences space (information depends on valid inference relative to a person's theory or epistemic state);
- the semantic approach (Floridi 2004a and 2004b) defines information in terms of data space (semantic information is well-formed, meaningful, and truthful data).

Each extentionalist approach can be given an intentionalist reading by interpreting the relevant space as a doxastic space, in which information is seen as a reduction in the degree of uncertainty or level of surprise given a state of knowledge of the informee (this is technically known as "interested information").

Communication theory approaches information as a physical phenomenon, syntactically. It is interested not in the usefulness, relevance, meaning, interpretation, or aboutness of data but in the level of detail and frequency in the uninterpreted data (signals or messages). It provides a successful mathematical theory because its central question is whether and how much data, not what information is conveyed. The other five approaches address the question "What is semantic information?" They seek to give an account of information as semantic content, usually adopting a propositional orientation (they analyze examples like "The cat is on the mat"). Does communication theory provide the necessary conditions for any theory of semantic information? Are semantic approaches mutually compatible? Is there a logical hierarchy? Do any of the previous approaches provide a clarification of the notion of data as well? Most of the problems in PI acquire a different meaning depending on how we answer this cluster of questions. Indeed, positions might be more compatible than they initially appear owing to different interpretations of the concept(s) of information involved.

Once the concept of information is clarified, each of the previous approaches needs to address the following question:

P.2: The I/O problem: What are the dynamics of information?

The question does not concern the nature of management processes (information seeking, data acquisition and mining, information harvesting and gathering, storage, retrieval, editing, formatting, aggregation, extrapolation, distribution, verification, quality control, evaluation, and so on); rather, it concerns information processes themselves, whatever goes on between the input and the output phase. Communication theory, as the mathematical theory of data transmission, provides the necessary conditions for any physical communication of information, but is otherwise of only marginal help. The information flow—understood as the carriage and transmission of information by some data about a referent, made possible by regularities in a distributed system—has been at the center of logical studies for some time (Barwise and Seligman 1997), but it still needs to be fully explored. How is it possible for something to carry information about something else? The problem here is not yet represented by the "aboutness" relation, which needs to be discussed in terms of meaning, reference and truth (see P.4 and P.5). The problem here concerns the nature of data as vehicles of information. In this version, the problem plays a central role in semiotics, hermeneutics, and situation logic. It is closely related to the problem of the naturalization of information. Various other logics, from classic first-order calculus to epistemic and erotetic logic, provide useful tools with which to analyze the logic of information (the logic of "S is informed that *p*"), but there is still much work to be done. For example, epistemic logic (the logic of "S knows that *p*") relies on a doxastic analysis of knowledge ("S believes that *p*"), and an open question is whether epistemic logic might be a fragment of information logic and the latter a fragment of doxastic logic. Likewise, recent approaches to the foundation of mathematics as a science of patterns (Resnik 2000) may turn out to provide enlightening insights into the dynamics of information, as well as benefiting from an approach in terms of information design (design seems to be a useful middle-ground concept between discovery and invention). Information processing, in the general sense of information-states transitions, includes at the moment effective computation (computationalism, Newell 1980; Pylyshyn 1984; Fodor 1975 and 1987; Dietrich 1990), distributed processing (connectionism, Smolensky 1988; Churchland and Sejnowski 1992), and dynamical-system processing (dynamism, Van Gelder 1995; Van Gelder and Port 1995; Eliasmith 1996). The relations among the current paradigms remain to be clarified (Minsky 1990, for example, argues in favor of a combination of computationalism and connectionism in AI, as does Harnad 1990 in cognitive science), as do the specific advantages and disadvantages of each, and the question as to whether they provide complete coverage of all possible *internalist* information-processing methods. I shall return to this point when discussing problems in AI.

The two previous questions are closely related to a third, more general problem:

P.3: The UTI challenge: Is a grand unified theory of information possible?

The reductionist approach holds that we can extract what is essential to understanding the concept of information and its dynamics from the wide variety of models, theories, and explanations proposed. The nonreductionist argues that we are probably facing a network of logically interdependent but mutually irreducible concepts. The plausibility of each approach needs to be investigated in detail. I personally side with Shannon and the nonreductionist. Both approaches, as well as any other solution in between, are confronted by the difficulty of clarifying how the various meanings of information are related, and whether some concepts of information are more central or fundamental than others and should be privileged. Waving a Wittgensteinian suggestion of family resemblance means acknowledging the problem, not solving it.

4. Semantics

> Evans had the idea that there is a much cruder and more fundamental concept than that of knowledge on which philosophers have concentrated so much, namely the concept of information. Information is conveyed by perception, and retained by memory, though also transmitted by means of language. One needs to concentrate on that concept before one approaches that of knowledge in the proper sense. Information is acquired, for example, without one's necessarily having a grasp of the proposition which embodies it; the flow of information operates at a much more basic level than the acquisition and transmission of knowledge. I think that this conception deserves to be explored. It's not one that ever occurred to me before I read Evans, but it is probably fruitful. It also distinguishes this work very sharply from traditional epistemology. [Dummett 1993, 186]

We have seen that most theories concentrate on the analysis of semantic information. Since much of contemporary philosophy is essentially philosophical semantics (a sort of theology without God), it is useful to carry on our review of problem areas by addressing now the cluster of issues arising in informational semantics. Their discussion is bound to be deeply influential in several areas of philosophical research. But first, a warning. It is hard to formulate problems clearly and in some detail in a completely theory-neutral way. So in what follows I have relied on the semantic frame, namely, the view that semantic information can be satisfactorily analyzed in terms of well-formed, meaningful, and truthful data. This semantic approach is simple and powerful enough for the task at hand. If the problems selected are sufficiently robust, it is reasonable to expect that their general nature and significance are not relative to the theoretical vocabulary in which they are cast but will be exportable across conceptual platforms.

We have already encountered the issue of the nature of data, in P.1. Suppose data are intuitively described as uninterpreted differences (symbols or signals). How do they become meaningful? This is

P.4: DGP, the data-grounding problem: How can data acquire their meaning?

Searle (1980) refers to a specific version of the data-grounding problem as the problem of intrinsic meaning or "intentionality." Harnad (1990) defines it as the symbols-grounding problem and unpacks it thus:

> How can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads? How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols?

Arguably, the frame problem (how a situated agent can represent, and interact with, a changing world satisfactorily) and its subproblems are a consequence of the data-grounding problem (Harnad 1994). We shall see that the data-grounding problem acquires a crucial importance in the artificial versus natural intelligence debate (see P.8–P.10). In more metaphysical terms, this is the problem of the semanticization of being, and it is further connected with the problem of whether information can be naturalized (see P.16). Can PI explain how the mind conceptualizes reality? (Mingers 1997).

Once grounded, meaningful data can acquire different truth values; the question is how:

P.5: The problem of alethization: How can meaningful data acquire their truth values?

P.4 and P.5 gain a new dimension when asked within epistemology and the philosophy of science, as we shall see in P.13 and P.14. They also interact substantially with the way in which we approach both a theory of truth and a theory of meaning, especially a truth-functional one. Are truth and meaning understandable on the basis of an informational approach, or is it information that needs to be analyzed in terms of noninformational theories of meaning and truth? To call attention to this important set of issues, it is worth asking two more place-holder questions:

P.6: Informational truth theory: Can information explain truth?

In this, as in the following question, we are not asking whether a specific theory could be couched, more or less metaphorically, in some informational vocabulary. This would be a pointless exercise. What is in question is not even the mere possibility of an informational approach. Rather, we are asking (a) whether an informational theory could explain truth more satisfactorily than other current approaches (Kirkham 1992) and (b) should (a) be answered in the negative, whether an informational approach could at least help to clarify the theoretical constraints to be satisfied by other approaches. Note that P.6 is connected with the information circle (P.12) and the possibility of an information view of science (P.14). The next question is:

P.7: Informational semantics: Can information explain meaning?

Several informational approaches to semantics have been investigated in epistemology (Dretske 1981 and 1988), situation semantics (Seligman and Moss 1997), discourse-representation theory (Kamp 1981), and dynamic semantics (Muskens et al. 1997). Is it possible to analyze meaning not truth-functionally but as the potential to change the informational context? Can semantic phenomena be explained as aspects of the empirical world? Since P.7 asks whether meaning can at least partly be grounded in an objective, mind- and language-independent notion of information (naturalization of intentionality), it is strictly connected with P.16, the problem of the naturalization of information.

5. Intelligence

> A computer program capable of acting intelligently in the world must have a general representation of the world in terms of which its inputs are interpreted. Designing such a program requires commitments about what knowledge is and how it is obtained. Thus, some of the major traditional problems of philosophy arise in artificial intelligence. [McCarthy and Hayes 1969]

Information and its dynamics are central to the foundations of AI, cognitive science, epistemology, and philosophy of science. Let us concentrate on the initial two first.

AI and cognitive science study cognitive agents as informational systems that receive, store, retrieve, transform, generate, and transmit information. This is the *information-processing* view. Before the development of connectionist and dynamic-system models of information processing, it was also known as the computational view. The latter expression was acceptable when a Turing machine (Turing 1936) and the machine involved in the Turing test (Turing 1950) were inevitably the same. It has become misleading, however, because computation, when used as a technical term (effective computation), refers now to the specific class of algorithmic symbolic processes that can be performed by a Turing machine, that is, recursive functions (Turing 1936; Minsky 1967; Boolos and Jeffrey 1989; Floridi 1999).

The information-processing view of cognition, intelligence, and mind provides the oldest and best-known cluster of significant problems in PI.[3] Some of their formulations, however, have long been regarded as uninteresting. Turing (1950) considered "Can machines think?" a meaningless way of posing the otherwise interesting problem of the functional differences between AI and NI (natural intelligence). Searle (1990) has equally dismissed "Is the brain a digital computer?" as ill defined. The same holds true of the unqualified question "Are naturally intelligent systems information-processing systems?" Such questions are vacuous. Informational concepts are so powerful that, given the right level of abstraction (LoA) (Floridi and Sanders 2004), anything can be presented as an information system, from a building to a volcano, from a forest to a dinner, from a brain to a company, and any process can be simulated informationally—heating, flying, and knitting. So pancomputationalist views have the hard task of providing a credible answer to the question of what it would mean for a physical system *not* to be an informational system (that is, a computational system, if computation is used to mean information processing, see Chalmers online and 1996). The task is hard because pancomputationalism does not seem vulnerable to a refutation, in the form of a realistic token counterexample in a world nomically identical to the one to which pancomputationalism is applied.[4] A good way of posing the problem is not: "Is '*x* is *y*' adequate?" but rather "If '*x* is *y*' at LoA *z*, is *z* adequate?" In what follows, I have

[3] In 1964, introducing his influential anthology, Anderson wrote that the field of philosophy of AI had already produced more than a thousand articles (Anderson 1964, 1). No wonder that (sometimes overlapping) editorial projects have flourished. Among the available titles, the reader may wish to keep in mind Ringle 1979 and Boden 1990, which provide two further good collections of essays, and Haugeland 1981,which was expressly meant to be a sequel to Anderson 1964 and was further revised in Haugeland 1997.

[4] Chalmers (online) seems to believe that pancomputationalism is empirically falsifiable, but what he offers is not (a) a specification of what would count as an instance of *x* that would show how *x* is not to be qualified computationally (or information-theoretically, in the language of this article) given the nomic characterization N of the universe, but rather (b) just a rewording of the idea that pancomputationalism might be false, i.e., a negation of the nomic characterization N of the universe in question: "To be sure, there are some ways that empirical science might prove it to be false: if it turns out that the fundamental laws of physics are noncomputable and if this noncomputability reflects itself in cognitive functioning, for instance, or if it turns out that our cognitive capacities depend essentially on infinite precision in certain analog quantities, or indeed if it turns out that cognition is mediated by some non-physical substance whose workings are not computable." To put it simply, we would like to be told something along the lines that a white raven would falsify the statement that all ravens are black, but instead we are told that the absence of blackness or of ravens altogether would, which it does not.

distinguished between problems concerning cognition and problems concerning intelligence.

A central question in cognitive science is:

P.8: Descartes' problem: Can (forms of) cognition C be fully and satisfactorily analyzed in terms of (forms of) information processing IP at some level of abstraction LoA? How is the triad ⟨C, IP, LoA⟩ to be interpreted?

The stress is usually on the types of C and IP involved and their mutual relations, but the LoA adopted and its level of adequacy play a crucial role (Marr 1982; Dennett 1994; McClamrock 1991). A specific LoA is adequate in terms of constraints and requirements. We need to ask first whether the analysis respects the constraints embedded in the selected observables we wish to model (for example: C is a dynamic process, but we have developed a static model). We then need to make sure that the analysis satisfies the requirements orienting the modeling process. Requirements can be of four general types: explanation (from the merely metaphorical to the fully scientific level), control (monitoring, simulating, or managing *x*'s behavior), modification (purposeful change of *x*'s behavior itself, not of its model), and construction (implementation or reproduction of *x* itself). We usually assume that LoAs come in a scale of granularity or detail, from higher (coarser-grained) to lower (finer-grained) levels, but this is not necessarily true if we concentrate on the requirements they satisfy. Consider a building. One LoA describes it in terms of architectural design, say as a Victorian house, another describes it in terms of property-market valuation, and a third describes it as Mary's house. A given LoA might be sufficient to provide an explanatory model of *x* without providing the means to implement *x*, and vice versa.

Answers to P.8 determine our orientation toward other specific questions (see Chalmers online) like: Is information processing sufficient for cognition? If it is, what is the precise relation between information processing and cognition? What is the relation between different sorts and theories of information processing, such as computationalism, connectionism, and dynamicism (Van Gelder and Port 1995; Van Gelder 1995; Garson 1996) for the interpretation of ⟨C, IP, LoA⟩? What are the sufficient conditions under which a physical system implements given information processing? For example, externalist or antirepresentationist positions stress the importance of "environmental," "situated" or "embodied" cognition (see Gibson 1979; Varela et al. 1991; Clancey 1997). Note that asking whether cognition is computable is not yet asking whether cognition is computation: *x* might be computable without necessarily being carried out computationally (Rapaport 1998).

The next two open problems (Turing 1950) concern intelligence in general rather than cognition in particular, and are central in AI:

P.9: The reengineering problem (Dennett 1994): Can (forms of) natural intelligence NI be fully and satisfactorily analyzed in terms of (forms of) information processing IP at some level of abstraction LoA? How is the triad ⟨NI, IP, LoA⟩ to be interpreted?

P.9 asks what kind or form of intelligence is being analyzed, what notion(s) of information is (are) at work here, which model of information dynamics correctly describes natural intelligence, what the level of abstraction adopted is, and whether it is adequate. For example, one could try an impoverished Turing test in which situated intelligent behavior, rather than purely dialogical interaction, is being analyzed by observing two agents, one natural and the other artificial, interacting with a problem environment modifiable by the observer (Harnad 2001). Imagine a robot and a cat searching for food in a maze: Would the observer placed in a different room be able to discriminate between the natural and the artificial agent? All this is not yet asking

P.10: Turing's problem: Can (forms of) natural intelligence be fully and satisfactorily implemented nonbiologically?

The question leaves open the possibility that NI might be an IP sui generis (Searle 1980) or just so complex as to elude forever any engineering attempt to duplicate it (Dreyfus 1992; Lucas 1961 and 1996; Penrose 1989, 1990, and 1994). Suppose, on the other hand, that NI is not, or is only incompletely, implementable nonbiologically, what is missing? Consciousness? Creativity? Freedom? Embodiment? All or perhaps some of these factors and even more? Alternatively, is it just a matter of the size, detail, and complexity of the problem? Even if NI is not implementable, is NI behavioral output still (at least partly) reproducible in terms of delivered effects by some implementable forms of information processing? These questions lead to a reformulation of "the father of all problems" (its paternity usually being attributed to Descartes) in the study of intelligence and the philosophy of mind:

P.11: The MIB (mind-information-body) problem: Can an informational approach solve the mind-body problem?

As usual, the problem is not about conceptual vocabulary or the mere possibility of an informational approach. Rather, we are asking whether an informational theory can help us to solve the difficulties faced by monist and dualist approaches. In this context, one could ask whether personal identity, for example, might be properly understood not in physical or mental terms but in terms of information space.

We can now move on to a different set of issues, concerning intelligence as the source of knowledge in epistemology and philosophy of science. The next cluster of problems requires a brief premise.

One of the major dissimilarities between current-generation artificial-intelligence systems (AIs) and human natural intelligences (NIs) is that AIs can identify and process only data (uninterpreted patterns of differences and invariances), whereas NIs can identify and process mainly information (in the weak sense of well-formed patterns of meaningful data). In saying that AIs are data systems whereas NIs are information systems, one should carefully avoid denying five things:

1. Young NIs, for example the young Augustine, seem to go through a formative process in which, at some stage, they experience only data, not information. Infants are information virgins;
2. adult NIs, for example the adult John Searle or a medieval copyist, could behave or be used as if they were perceiving only data, not information. One could behave like a child—or an Intel processor —if one is placed in a Chinese room or, more realistically, is copying a Greek manuscript without knowing even the alphabet of the language, just the physical shape of the letters;
3. cognitively, psychologically, or mentally impaired NIs, including the old Nietzsche, might also act like children and fail to experience information (like "this is a horse") when exposed to data;
4. there is certainly a neurochemical level at which NIs process data, not yet information;
5. NIs' semantic constraints might be comparable to, or even causally connected with, AIs' syntactic constraints, at some adequate LoA.

Fully and normally developed NIs seem entrapped in a semantic stance. Strictly speaking, we do not consciously cognize pure meaningless data. What goes under the name of "raw data" are data that might lack a specific and relevant interpretation, not any interpretation (this is true even for John Searle and the medieval copyist: one sees Chinese characters, the other Greek letters, although they do not know that this is what the characters are). The genuine perception of completely uninterpreted data might be possible under very special circumstances, but it is not the norm and cannot be part of a continuously sustainable, conscious experience, at least because we never perceive data in isolation but always in a semantic context that attributes some meaning to them (it does not have to be the right meaning, as John Searle and the medieval copyist show). Usually, when human NIs seem to perceive data, this is only because they are used to dealing with such rich semantic contents that they mistake dramatically impoverished or variously interpretable information for something completely devoid of any semantic content. On the other hand, computers are often and rightly described as purely syntactic machines, yet "purely syntactic" is a comparative abstraction, like "virtually fat free." It means that the level of semantic stance is negligible, not that it is completely nonexistent. Computers are capable of (responding to) elementary discrimination (the detection of an identity as an identity and of a difference not in terms of perception of the peculiar and rich features of the entities involved but as a simple registration of an invariant lack of identity constituting the relata as relata), and this is after all a protosemantic act. Unfortunately, discrimination is far too poor to generate anything resembling a semantic stance and suffices only to guarantee an efficient manipulation of discrimination-friendly data. It is also the only vaguely protosemantic act that (present) computers are able to perform as "cognitive systems," the rest being extrinsic semantics, simulated only through syntax, prerecorded memory, layers of interfaces, and human-computer interaction (HCI).

Thus, at the moment, data as interpretable but uninterpreted and discriminable differences represent the semantic upper limit of AIs but the semantic lower limit of NIs, which normally deal with information. Ingenious layers of interfaces exploit this threshold and make possible HCI. The specification indicates that current AI achievements are constrained by syntactical resources, whereas NI achievements are constrained by semantic ones. To understand the informational stance as a constraint, one only needs to consider any nonnaive epistemology. Kant's dichotomy between noumena and phenomena, for example, could be interpreted as a dichotomy between data and information, with the *Umwelt* of experience as the threshold where the flow of uninterpreted data regularly and continuously collapses into information flow. Note that conceding some minimal protosemantic capacity to a computer works in favor of an extensionalist conception of information as being "in the world" rather than just in the mind of the informee. I shall return to this issue when discussing P.16.

We are now ready to appreciate a new series of questions:

P.12: The informational circle: How can information be audited? If information cannot be transcended but can only be checked against further information—if it is information all the way up and all the way down—what does this tell us about our knowledge of the world?

The informational circle is reminiscent of the hermeneutical circle. It underpins the modern debate on the foundation of epistemology and the acceptability of some form of realism in the philosophy of science, according to which our information about the world captures something of the way the world is (Floridi 1996). It is closely related both to P.6 and to the next two questions:

P.13: The continuum hypothesis: Should epistemology be based on a theory of information?

Knowledge is often said to presuppose information in the light of a "continuum hypothesis" that knowledge encapsulates truth because it encapsulates semantic information (see P.5). Compared to information, knowledge is a rare phenomenon indeed. Even in a world without Gettier-like tricks, we must confess to being merely informed about most of what we think we know, if knowing demands being able to provide a sound explanation or a justification of what one is informed about. Before answering P.13, however, one should also consider that some theories of information, for example, internalist or intentionalist approaches, interpret information as depending upon knowledge, not vice versa. Can there be information states without epistemic states (see P.15–P.16)? What is knowledge from an information-based approach? If knowledge does presuppose information, could this help to solve Gettier-type problems? (In Floridi 2004d I argue that it does, by showing that the Gettier problem cannot be solved). Is it possible that (1) S has the true belief that *p* and yet (2) S is not informed that *p*? (Barwise and Seligman [1997, 9] hold it is.) These questions have been addressed by information-theoretic epistemologists for some time now, but they still need to be fully investigated. When it comes to scientific knowledge, it seems that the value of an informational turn can be stressed by investigating the following question:

P.14: The semantic view of science: Is science reducible to information modeling?

In some contexts (probability or modal states and inferential spaces), we adopt a conditional, laboratory view. We analyze what happens in "*a*'s being (of type, or in state) *F* is correlated to *b* being (of type, or in state) *G*, thus carrying for the observer the information that *b* is *G*" (Barwise and Seligman [1997] provide a similar analysis based on Dretske 1981) by assuming that *F*(*a*) and *G*(*b*). In other words, we assume a given model. The question asked here is: How do we build the original model? Many approaches seem to be ontologically overcommitted. Instead of assuming a world of empirical affordances and constraints to be designed, they assume a world already well modeled, ready to be discovered. The semantic approach to scientific theories (Suppes 1960 and 1962; van Fraassen 1980; Giere 1988; Suppe 1989), on the other hand, argues that "scientific reasoning is to a large extent model-based reasoning. It is models almost all the way up and models almost all the way down" (Giere 1999, 56). Theories do not make contact with phenomena directly; rather, higher models are brought into contact with other, lower models. These are themselves theoretical conceptualizations of empirical systems, which constitute an object being modeled as an object of scientific research. Giere (1988) takes most scientific models of interest to be nonlinguistic abstract objects. Models, however, are the medium, not the message. Is information the (possibly nonlinguistic) content of these models? How are informational models (semantically, cognitively, and instrumentally) related to the conceptualizations that constitute their empirical references? What is their semiotic status—for example, structurally homomorphic or isomorphic representations or data-driven and data-constrained informational constructs? What levels of abstraction are involved? Is science a social (multi-agent), information-designing activity? Is it possible to import, in (the philosophy of) science, modeling methodologies devised in information-system theory? Can an informational view help to bridge the gap between science and cognition? Answers to these questions are closely connected with the discussion of the problem of an informational theory of truth (P.6) and of meaning (P.7).

The possibility of a more or less informationally constructionist philosophy of science leads to our next cluster of problems, concerning the relation between information and the natural world.

6. Nature

> If the world were a completely chaotic, unpredictable affair, there would be no information to process. Still, the place of information in the natural world of biological and physical systems is far from clear. [Barwise and Seligman 1997, xi]

This lack of clarity prompts three families of problems.

P.15: Wiener's problem: What is the ontological status of information?

Most people agree that there is no information without (data) representation. Following Landauer and Bennett 1985 and Landauer 1987, 1991, and 1996, this principle is often interpreted materialistically, as advocating the impossibility of physically disembodied information, through the equation "representation = physical implementation." The view that there is no information without physical implementation is an inevitable assumption when working on the physics of computation, since computer science must necessarily take into account the physical properties and limits of the carriers of information. It is also the ontological assumption behind the Physical Symbol System Hypothesis in AI and cognitive science (Newell and Simon 1976). However, the fact that information requires a representation does not seem to entail that the latter ought to be physically implemented. Arguably, environments in which there are only noetic entities, properties, and processes (for example, Berkeley and Spinoza), or in which the material or extended universe has a noetic or nonextended matrix as its ontological foundation (for example, Pythagoras, Plato, Leibniz, and Hegel), seem perfectly capable of upholding the representationalist principle without also embracing a materialist interpretation. The relata giving rise to information could be monads, for example. So the problem here becomes: Is the informational an independent ontological category, different from the physical/material and (assuming one could draw this Cartesian distinction) the mental? Wiener, for example, thought that "information is information, not matter or energy. No materialism which does not admit this can survive at the present day" (1948, 132). If the informational is not an independent ontological category, to which category is it reducible? If it is an independent ontological category, how is it related to the physical/material and the mental? Answers to these questions determine the orientation a theory takes with respect to the following problem:

P.16: The problem of localization: Can information be naturalized?

The problem is connected with P.4—namely, the semanticization of data. It seems hard to deny that information is a natural phenomenon, so this is not what one should be asking here. Even elementary forms of life, such as sunflowers, survive only because they are capable of informational processes. The problem here is whether there is information in the world independently of forms of life capable of extracting it and if so, what kind of information is in question (an informational version of the teleological argument for the existence of God argues both that information is a natural phenomenon and that the occurrence of environmental information requires an intelligent source). If the world is sufficiently information rich, perhaps an agent may interact successfully with it by using "environmental information" directly, without being forced to go through a representation stage in which the world is first analyzed informationally. "Environmental information" still presupposes (or perhaps is identical with) some physical support, but it does not require any higher-level cognitive representation or computational processing to be immediately usable. This is argued, for example, by researchers in AI working on animats (artificial animals, either computer simulated or robotic). Animats are simple reactive agents, stimulus driven. They are capable of elementary, "intelligent" behavior, despite the fact that their design excludes the possibility of internal representations of the environment and any effective computation (see Mandik 2002 for an overview; the case for nonrepresentational intelligence is famously made in Brooks 1991).

So, are cognitive processes continuous with processes in the environment? Is semantic content (at least partly) external (Putnam)? Does "natural" or "environmental" information pivot on natural signs (Peirce) or on nomic regularities? Consider the typical example provided by the concentric rings visible in the wood of a cut tree trunk, which may be used to estimate the age of the plant. The externalist/extensionalist, who favors a positive answer to P.16 (for example, Dretske or Barwise), is faced by the difficulty of explaining what kind of information and how much of it saturates the world, what kind of access to, or interaction with, "information in the world" an informational agent can enjoy, and how information dynamics is possible. The internalist/intentionalist (for example, Fodor or Searle), who privileges a negative answer to P.16, needs to explain in what specific sense information depends on intelligence and whether this leads to an antirealist view.

The location of information is related to the question of whether there can be information without an informee, or whether information, in at least some crucial sense of the word, is essentially parasitic on the meanings in the mind of the informee, and the most it can achieve, in terms of ontological independence, is systematic interpretability. Before the discovery of the Rosetta stone, was it legitimate to regard Egyptian hieroglyphics as information, even if their semantics was beyond the comprehension of any interpreter? I've already mentioned that admitting that computers perform some minimal level of protosemantic activity works in favor of a "realist" position about "information in the world." Before moving to the next problem, it remains to be clarified whether the previous two ways of locating information might not be restrictive. Could information be neither here (intelligence) nor there (natural world) but on the threshold, as it were, as a special relation or interface between the world and its intelligent inhabitants (constructionism)? Or could it even be elsewhere, in a third world, intellectually accessible by intelligent beings but not ontologically dependent on them (Platonism)?

P.17: The It from Bit hypothesis (Wheeler 1990): Can nature be informationalized?

The neologism *informationalized* is ugly but useful to point out that this is the converse of the previous problem. Here too, it is important to clarify what the problem is not. We are not asking whether the metaphorical interpretation of the universe as a computer is more useful than misleading. We are not even asking whether an informational description of the universe, as we know it, is possible, at least partly and piecemeal. This is a challenging task, but formal ontologies already provide a promising answer (Smith 2003). We are asking whether the universe in itself could essentially be made of information, with natural processes, including causation, as special cases of information dynamics (for example, information flow and algorithmic, distributed computation, and forms of emergent computation). Depending on how one approaches the concept of information, it might be necessary to refine the problem in terms of digital data or other informational notions.

Answers to P.17 deeply affect our understanding of the distinction between virtual and material reality, of the meaning of artificial life in the ALife sense (Bedau 2003), and of the relation between the philosophy of information and the foundations of physics: If the universe is made of information, is quantum physics a theory of physical information? Moreover, does this explain some of its paradoxes? If nature can be informationalized, does this help to explain how life emerges from matter, and hence how intelligence emerges from life? Of course these questions are closely related to questions listed in section 5: "Can we build a gradualist bridge from simple amoeba-like automata to highly purposive intentional systems, with identifiable goals, beliefs, etc.?" (Dennett 1998, 262).

7. Values

> It has long been clear to me that the modern ultra-rapid computing machine was in principle an ideal central nervous system to an apparatus for automatic control; and that its input and output need not be in the form of numbers or diagrams but might very well be, respectively, the readings of artificial sense organs, such as photoelectric cells or thermometers, and the performance of motors or solenoids [. . .] we are already in a position to construct artificial machines of almost any degree of elaborateness of performance. Long before Nagasaki and the public awareness of the atomic bomb, it had occurred to me that we were here in the presence of another social potentiality of unheard-of importance for good and for evil. [Wiener 1948, 27–28]

The impact of ICT on contemporary society has caused new and largely unanticipated ethical problems (see Bynum 1998 and Johnson 2000 for an overview). In order to fill this policy and conceptual vacuum (Moor 1985), computer ethics (CE) carries out an extended and intensive study of real-world issues, usually in terms of reasoning by analogy. At least since the 1970s (see Bynum 2000 for earlier works in CE), CE's focus has moved from problem analysis—primarily aimed at sensitizing public opinion, professionals, and politicians—to tactical solutions resulting, for example, in the evolution of professional codes of conduct, technical standards, usage regulations, and new legislation. The constant risk of this bottom-up procedure has remained the spreading of ad hoc or casuistic approaches to ethical problems. Prompted partly by this difficulty, and partly by a natural process of self-conscious maturation as an independent discipline, CE has further combined tactical solutions with more strategic and global analyses. The "uniqueness debate" on the foundation of CE is an essential part of this top-down development (Floridi and Sanders 2002; Tavani 2002). It is characterized by a metatheoretical reflection on the nature and justification of CE and on whether the moral issues confronting CE are unique, and hence whether CE should be developed as an independent field of research with a specific area of application and an autonomous, theoretical foundation. The problem here is:

P.18: The uniqueness debate: Does computer ethics have a philosophical foundation?

Once again, the question is intentionally general. Answering it means addressing the following questions: Why does ICT raise moral issues?

Can CE amount to a coherent and cohesive discipline, rather than a more or less heterogeneous and random collection of ICT-related ethical problems, applied analyses, and practical solutions? If so, what is its conceptual rationale? How does it compare with other (applied) ethical theories? Are CE issues unique (in the sense of requiring their own theoretical investigations, not entirely derivative from standard ethics)? Alternatively, are they simply moral issues that happen to involve ICT? What kind of ethics is CE? What justifies a certain methodology in CE, for example, reasoning by analogy and case-based analysis? What is CE's rationale? What is the contribution of CE to the ethical discourse?

8. Conclusion

We have now come to the end of this review. I hope the reader will be thrilled rather than depressed by the amount of work that lies ahead. I must confess I find it difficult to provide an elegant way of closing this article. Since it analyzes questions but provides no answers, it should really end with "The Beginning" rather than "The End." However, as I relied on Hilbert to introduce the topic, I might as well quote him again to conclude it:

> To such a review of problems the present day, lying at the meeting of the centuries, seems to me well adapted. For the close of a great epoch not only invites us to look back into the past but also directs our thoughts to the unknown future.

Exactly.

Wolfson College
University of Oxford
Oxford OX2 6UD
UK
luciano.floridi@philosophy.oxford.ac.uk

Acknowledgments

This article is a revised version of the Herbert A. Simon Lecture on Computing and Philosophy I gave at CAP@CMU, the 2001 Computing and Philosophy Conference at Carnegie Mellon University. I wish to thank Robert Cavalier, president of the International Association for Computing and Philosophy (IACAP) and of the APA Committee on Philosophy and Computers, for the invitation to deliver the lecture, and many colleagues, especially Terry Bynum and Jim Moor, for their feedback during and after the conference.

References

Anderson, A. R. 1964. *Minds and Machines*. Englewood Cliffs: Prentice-Hall.

Bar-Hillel, Y., and R. Carnap. 1953. "An Outline of a Theory of Semantic Information." In Bar-Hillel 1964, 221–74.

Bar-Hillel, Y. 1964. *Language and Information*. Reading, Mass., and London: Addison Wesley.

Barwise, J., and J. Perry. 1983. *Situations and Attitudes*. Cambridge, Mass.: MIT Press.

Barwise, J., and J. Seligman. 1997. *Information Flow—The Logic of Distributed Systems*. Cambridge: Cambridge University Press.

Bedau, M. 2003. "Artificial Life." In Floridi 2004, 197–211.

Boden, M. A. (ed.) 1990. *The Philosophy of Artificial Intelligence*. Oxford Readings in Philosophy. Oxford: Oxford University Press.

Boolos, G., and R. Jeffrey. 1989. *Computability and Logic*. Third edition. Cambridge: Cambridge University Press.

Brooks, R. 1991. "Intelligence without Representation." *Artificial Intelligence* 47:139–59.

Burkholder, L. (ed.) 1992. *Philosophy and the Computer*. Boulder, San Francisco, and Oxford: Westview Press.

Bynum, T. W. 1998. "Global Information Ethics and the Information Revolution." In Bynum and Moor 1998, 274–89.

———. 2000. "A Very Short History of Computer Ethics." *APA Newsletter on Philosophy and Computers* 99, no. 2 (spring). http://www.southernct.edu/organizations/rccs/resources/research/introduction/bynum_shrt_hist.html.

Bynum, T. W., and J. H. Moor (eds.). 1998. *The Digital Phoenix: How Computers Are Changing Philosophy*. Oxford: Blackwell.

Chalmers, D. 1996. "Does a Rock Implement Every Finite-State Automaton?" *Synthese* 108:309–33. http://www.u.arizona.edu/~chalmers/papers/rock.html.

———. Online. "A Computational Foundation for the Study of Cognition." http://www.u.arizona.edu/~chalmers/papers/computation.html.

Churchland, P. S., and T. J. Sejnowski. 1992. *The Computational Brain*. Cambridge, Mass.: Bradford/MIT Press.

Clancey, W. J. 1997. *Situated Cognition*. Cambridge: Cambridge University Press.

Colburn, T. R. 2000. *Philosophy and Computer Science*. Armonk, N.Y., and London: M. E. Sharpe.

Dennett, D. C. 1994. "Cognitive Science as Reverse Engineering: Several Meanings of 'Top-Down' and 'Bottom-Up.'" In *Logic, Methodology, and Philosophy of Science IX*, edited by D. Prawitz, B. Skyrms, and D. Westerstahl, 679–89. Amsterdam: Elsevier Science. http://cogsci.soton.ac.uk/~harnad/Papers/Py104/dennett.eng.html.

———. 1998. *Brainchildren*. Cambridge, Mass.: MIT Press.

Devlin, K. 1991. *Logic and Information*. Cambridge: Cambridge University Press.

Dietrich, E. 1990. "Computationalism." *Social Epistemology* 4:135–54.

Dretske, F. 1981. *Knowledge and the Flow of Information*. Cambridge, Mass.: MIT Press. Reprint Stanford: CSLI, 1999.

———. 1988. *Explaining Behavior*. Cambridge, Mass.: MIT Press.

Dreyfus, H. L. 1992. *What Computers Still Can't Do—A Critique of Artificial Intelligence*. Second edition. Cambridge, Mass.: MIT Press.

Dummett, M. 1993. *Origins of Analytical Philosophy*. London: Duckworth.

Eliasmith, C. 1996. "The Third Contender: A Critical Examination of the Dynamicist Theory of Cognition." *Journal of Philosophical Psychology* 9, no. 4:441–63.

Floridi, L. 1996. *Scepticism and the Foundation of Epistemology—A Study in the Metalogical Fallacies*. Leiden: Brill.

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

———. 2003. "What Is the Philosophy of Information?" In Moor and Bynum 2002, 117–38.

——— (ed.) 2004. *The Blackwell Guide to the Philosophy of Computing and Information*. Oxford: Blackwell.

———. 2004a. "Outline of a Theory of Strongly Semantic Information." *Minds and Machines*, forthcoming. Preprint available at http://www.wolfson.ox.ac.uk/~floridi/pdf/otssi.pdf.

———. 2004b. "Is Information Meaningful Data?" *Philosophy and Phenomenological Research*, forthcoming. Preprint available at http://www.wolfson.ox.ac.uk/~floridi/pdf/iimd.pdf.

———. 2004c. "Information, Semantic Conceptions of." Entry in the *Stanford Encyclopedia of Philosophy*, forthcoming.

———. 2004d. "On the Logical Insolvability of the Gettier Problem." *Synthese*, forthcoming. Preprint available at http://www.wolfson.ox.ac.uk/~floridi/pdf/oligp.pdf.

Floridi, L., and J. W. Sanders. 2002. "Computer Ethics: Mapping the Foundationalist Debate." *Ethics and Information Technology* 4, no. 1:1–9.

———. 2004. "The Method of Abstraction." In *Yearbook of the Artificial: Nature, Culture, and Technology: Models in Contemporary Sciences*, edited by M. Negrotti. Bern: Peter Lang. Preprint available at http://www.wolfson.ox.ac.uk/~floridi/pdf/tmola.pdf.

Fodor, J. A. 1975. *The Language of Thought*. New York: Thomas Y. Crowell.

———. 1987. *Psychosemantics*. Cambridge, Mass.: MIT/Bradford.

Garson, J. W. 1996. "Cognition Poised at the Edge of Chaos: A Complex Alternative to a Symbolic Mind." *Philosophical Psychology* 9: 301–22.

Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Boston, Mass.: Houghton Mifflin.

Giere, R. N. 1988. *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

———. 1999. "Using Models to Represent Reality." In *Model-Based Reasoning in Scientific Discovery*, edited by L. Magnani, N. J. Nersessian, and P. Thagard. Dordrecht: Kluwer.

Grim, P., G. Mar, and P. St. Denis. 1998. *The Philosophical Computer*. Cambridge, Mass.: MIT Press.

Groenendijk, J., and M. Stokhof. 1991. "Dynamic Predicate Logic." *Linguistics and Philosophy* 14:39–100.

Harnad, S. 1990. "The Symbol Grounding Problem." *Physica* D 42:335–46. http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad90.sgproblem.html.

———. 1993. "Problems, Problems: The Frame Problem as a Symptom of the Symbol Grounding Problem." *Psycoloquy* 4:34: http://cogsci.soton.ac.uk/harnad/Papers/Harnad/harnad93.frameproblem.html.

———. 2001. "Minds, Machines, and Turing: The Indistinguishability of Indistinguishables." *Journal of Logic, Language, and Information* 9, no. 4:425–45. Special issue entitled *Alan Turing and Artificial Intelligence*. http://www.wkap.nl/journalhome.htm/0925-8531.

Harris, R. A. 1998. "A Note on the Max Planck Effect." *Rhetoric Society Quarterly* 28:85–89.

Haugeland, J. 1981. *Mind Design: Philosophy, Psychology, Artificial Intelligence*. Montgomery, Vt.: Bradford Books.

———. 1997. *Mind Design II: Philosophy, Psychology, Artificial Intelligence*. Cambridge, Mass.: MIT Press.

Hilbert, D. 1900. "Mathematische Probleme: Vortrag, gehalten auf dem internationalen Mathematiker-Kongress zu Paris 1900." In *Nachrichten von der königlichen Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse, Geschäftliche Mitteilungen*, 253–97. Reprinted in *Archiv der Mathematik und Physik* 3, no. 1 (1901): 44–63, 213–37. English translation "Mathematical Problems." *Bulletin of the American Mathematical Society* 8 (1902): 437–79. http://aleph0.clarku.edu/~djoyce/hilbert/problems.html.

Israel, D., and J. Perry. 1990. "What Is Information?" In *Information, Language, and Cognition*, edited by P. Hanson, 1–19. Vancouver: University of British Columbia Press.

Johnson, D. 2000. *Computer Ethics*. Third edition. Upper Saddle River, N.J.: Prentice Hall.

Kamp, H. 1984. "A Theory of Truth and Semantic Interpretation." In *Truth, Interpretation, and Information*, edited by J. Groenendijk, T. M. V. Janssen, and M. Stokhof, 277–322. Dordrecht: Foris.

Kirkham, R. L. 1992. *Theories of Truth: A Critical Introduction*. Cambridge, Mass.: MIT Press.

Landauer, R. 1987. "Computation: A Fundamental Physical View." *Physica Scripta* 35:88–95.
———. 1991. "Information Is Physical." *Physics Today* 44:23–29.
———. 1996. "The Physical Nature of Information." *Physics Letter* A 217:188.
Landauer, R., and C. H. Bennett. 1985. "The Fundamental Physical Limits of Computation." *Scientific American* (July): 48–56.
Lucas, J. R. 1961. "Minds, Machines and Gödel." *Philosophy* 36: 112–27.
———. 1996. "Minds, Machines, and Gödel: A Retrospect." In *Machines and Thought: The Legacy of Alan Turing*, edited by P. Millican and A. Clark, 103–24. Oxford: Clarendon Press.
Mandik, P. 2002. "Synthetic Neuroethiology." In Bynum and Moor 2002, 11–29.
Marr, D. 1982. *Vision*. San Francisco: Freeman.
McCarthy, J., and P. J. Hayes. 1969. "Some Philosophical Problems from the Standpoint of Artificial Intelligence." *Machine Intelligence* 4: 463–502. http://www.formal.stanford.edu/jmc/mcchay69.html.
McClamrock, R. 1991. "Marr's Three Levels: A Reevaluation." *Minds and Machines* 1:185–6.
McClelland, J. L., and D. E. Rumelhart (eds.). 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, Mass.: MIT Press.
McDowell, J. 1994. *Mind and World*. Cambridge, Mass.: Harvard University Press.
Mingers, J. 1997. "The Nature of Information and Its Relationship to Meaning." In Winder et al. 1997, 73–84.
Minsky, M. L. 1967. *Computation: Finite and Infinite Machines*. Englewood Cliffs, N.J.: Prentice Hall.
———. 1990. "Logical vs. Analogical or Symbolic vs. Connectionist or Neat vs. Scruffy." In *Artificial Intelligence at MIT: Expanding Frontiers*, vol. 1, edited by P. H. Winston. Cambridge, Mass.: MIT Press.
Mitcham, C., and A. Huning (eds.). 1986. *Philosophy and Technology II: Information Technology and Computers in Theory and Practice*. Dordrecht and Boston: Reidel.
Moor, J. H. 1985. "What Is Computer Ethics?" *Metaphilosophy* 16, no. 4: 266–75. http://www.southernct.edu/organizations/rccs/resources/teaching/teaching_mono/moor/moor_definition.html.
Moor, J. H., and T. W. Bynum (eds.). 2002. *Cyberphilosophy: The Intersection of Philosophy and Computing*. Special issue of *Metaphilosophy* 33, nos. 1/2 (January). Oxford: Blackwell.
Muskens, R., et al. 1997. "Dynamics." In van Benthem and ter Meulen 1997, 587–648.
Newell, A. 1980. "Physical Symbol Systems." *Cognitive Science* 4: 135–83.

Newell, A., and H. A. Simon. 1976. "Computer Science as Empirical Inquiry: Symbols and Search." *Communications of the ACM* 19 (March): 113–26.

Penrose, R. 1989. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford: Oxford University Press.

———. 1990. "Précis of 'The Emperor's New Mind': Concerning Computers, Minds, and the Laws of Physics." *Behavioral and Brain Sciences* 13:643–705.

———. 1994. *Shadows of the Mind: A Search for the Missing Science of Consciousness*. New York: Oxford University Press.

Planck, M. 1950. *Scientific Autobiography and Other Papers*. London: Williams and Norgate.

Pylyshyn, Z. W. 1984. *Computation and Cognition*. Cambridge, Mass.: MIT/Bradford.

Rapaport, W. J. 1998. "How Minds Can Be Computational Systems." *Journal of Experimental and Theoretical Artificial Intelligence* 10: 403–19.

Resnik, M. D. 2000. *Mathematics as a Science of Patterns*. Oxford: Oxford University Press.

Ringle, M. 1979. *Philosophical Perspectives in Artificial Intelligence*. Atlantic Highlands, N.J.: Humanities Press.

Searle, J. R. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3, no. 3:417–57.

———. 1990. "Is the Brain a Digital Computer?" *Proceedings and Addresses of the American Philosophical Association* 64:21–37.

Seligman, J., and L. S. Moss. 1997. "Situation Theory." In van Benthem and ter Meulen 1997, 239–309.

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27:379–423, 623–56.

———. 1993. *Collected Papers*. Edited by N. J. A. Sloane and A. D. Wyner. Los Alamos, Calif.: IEEE Computer Society Press.

Shannon, C. E., and W. Weaver. 1949. *The Mathematical Theory of Communication*. Urbana, Ill.: University of Illinois Press.

Simon, H. A. 1962. "The Computer as a Laboratory for Epistemology." In Burkholder 1992, 3–23.

Smith, B. 2003. "Ontology." In Floridi (ed.) 2004, 155–66.

Smolensky, P. 1988. "On the Proper Treatment of Connectionism." *Behavioral and Brain Sciences* 11, no. 1:1–23.

Spice, B. 2000. "CMU's Simon reflects on how computers will continue to shape the world." *Post-Gazette Science* (October 16). http://www.post-gazette.com/regionstate/20001016simon2.asp.

Suppe, F. 1989. *The Semantic Conception of Theories and Scientific Realism*. Urbana, Ill.: University of Illinois Press.

Suppes, P. 1960. "A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences." *Synthese* 12:287–301.

———. 1962. “Models of Data.” In *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress*, edited by E. Nagel, P. Suppes, and A. Tarski, 252–61. Stanford: Stanford University Press.

Tavani, H. T. 2002. “The Uniqueness Debate in Computer Ethics: What Exactly Is at Issue, and Why Does It Matter?” *Ethics and Information Technology* 4, no. 1:37–54.

Turing, A. M. 1936. “On Computable Numbers, with an Application to the Entscheidungsproblem.” *Proceedings of the London Mathematics Society*, 2d series, 42:230–65. Correction published in 43 (1936): 544–46.

———. 1950. “Computing Machinery and Intelligence.” *Mind* 59:236, 433–60.

van Fraassen, B. 1980. *The Scientific Image*. Oxford: Clarendon Press.

van Gelder, T. 1995. “What Might Cognition Be, if Not Computation?” *Journal of Philosophy* 92:345–81.

van Gelder, T., and R. Port (eds.). 1995. *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, Mass.: MIT Press.

van Benthem, J., and A. ter Meulen (eds.). 1997. *Handbook of Logic and Language*. Amsterdam: Elsevier.

Varela, F. J., E. Thompson, and E. Rosch. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, Mass.: MIT Press.

Wheeler, J. A. 1990. “Information, Physics, Quantum: The Search for Links.” In *Complexity, Entropy, and the Physics of Information*, edited by W. H. Zureck. Redwood City, Calif.: Addison Wesley.

Wiener, N. 1961. *Cybernetics or Control and Communication in the Animal and the Machine*. Second edition. Cambridge, Mass.: MIT Press Originally published in 1948.

Winder, R. L., S. K. Probert, and I. A. Beeson. 1997. *Philosophical Aspects of Information Systems*. London: Taylor and Francis.