

Section I

Introduction

What is the Philosophy of Information?

Luciano Floridi

1. Introduction: Philosophy of AI as a Premature Paradigm of PI

Andre Gide once wrote that one does not discover new lands without consenting to lose sight of the shore for a very long time. Looking for new lands, in 1978 Aaron Sloman heralded a new AI-based paradigm in philosophy. In a book appropriately entitled *The Computer Revolution in Philosophy*, he conjectured

1. that within a few years, if there remain any philosophers who are not familiar with some of the main developments in artificial intelligence, it will be fair to accuse them of professional incompetence, and
2. that to teach courses in philosophy of mind, epistemology, aesthetics, philosophy of science, philosophy of language, ethics, metaphysics and other main areas of philosophy, without discussing the relevant aspects of artificial intelligence will be as irresponsible as giving a degree course in physics which includes no quantum theory. (Sloman 1978, p. 5, numbered structure added).

Sloman was not alone. Other researchers before and after him (Simon 1962; McCarthy and Hayes 1969; McCarthy 1995; Pagels 1988, who argues in favour of a complexity theory paradigm; Burkholder 1992, who speaks of a “computational turn”) correctly perceived that the practical and conceptual transformations caused by ICS (Information and Computation Sciences) and ICT (digital Information and Communication Technologies) were bringing about a macroscopic change, not only in science, but in philosophy too. It was the so-called “computer revolution” or “information turn”. Their forecasts, however, underestimated the unrelenting difficulties that the acceptance of a new paradigm would encounter. Turing began publishing his seminal papers in the 1930s. During the following fifty years, cybernetics, information theory, AI, system theory, computer science, complexity theory and ICT attracted some significant but comparatively sporadic and marginal interest from the philosophical community, especially in terms of philosophy of AI. 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). Since then, editorial projects have flourished (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 as a sequel to Anderson

1964 and was further revised in Haugeland 1997). Work in the philosophy of AI prepared the ground for the emergence of an independent field of investigation and a new computational and information-theoretic approach in philosophy. Until the 1980s, however, the philosophy of AI failed to give rise to a mature, innovative and influential program of research, let alone a revolutionary change of the magnitude and importance envisaged by researchers like Sloman in the 1970s. With hindsight, it is easy to see how AI could be perceived as an exciting new field of research and the source of a radically innovative approach to traditional problems in philosophy.

Ever since Alan Turing's influential paper "Computing machinery and intelligence" [...] and the birth of the research field of Artificial Intelligence (AI) in the mid-1950s, there has been considerable interest among computer scientists in theorising about the mind. At the same time there has been a growing feeling amongst philosophers that the advent of computing has decisively modified philosophical debates, by proposing new theoretical positions to consider, or at least to rebut. (Torrance, 1984, p. 11)

The philosophy of AI acted as a Trojan horse, introducing a more encompassing computational/informational paradigm into the philosophical citadel (earlier statements of this view can be found in Simon 1962, Pylyshyn 1970, and Boden 1984; and more recently in McCarthy 1995, Sloman 1995 and Simon 1996). For reasons that will be clarified in section four, I suggest we refer to this new paradigm as PI, *philosophy of information*.

Until the mid-1980s, PI was still premature and perceived as transdisciplinary rather than interdisciplinary. The philosophical and scientific communities were not yet ready for it. The cultural and social contexts were equally unprepared. Each factor deserves a brief clarification.

Like other intellectual enterprises, PI deals with three types of domains: *topics* (facts, data, problems, phenomena, observations, etc.); *methods* (techniques, approaches, etc.); and *theories* (hypotheses, explanations, etc.). A discipline is *premature* if it attempts to innovate in more than one of these domains simultaneously, thus detaching itself too abruptly from the normal and continuous thread of evolution of its general field (Stent 1972). A quick look at the two points made by Sloman in his prediction shows that this was exactly what happened to PI in its earlier appearance as the philosophy of AI.

The inescapable interdisciplinarity of PI further hindered the prospects for a timely recognition of its significance. Even now, many philosophers seem content to consider many topics in PI to be worth the attention only of researchers in English,

Mass Media, Cultural Studies, Computer Science or Sociology Departments, to mention a few examples. PI needed philosophers accustomed to conversing with cultural and scientific issues across the boundaries, and these were not to be found easily. Too often, everyone's concern is nobody's business and, until the recent development of the information society, PI was perceived to be at too much of a crossroads of technical matters, theoretical issues, applied problems and conceptual analyses to be anyone's own area of specialisation. PI was perceived to be transdisciplinary like cybernetics or semiotics, rather than interdisciplinary like biochemistry or cognitive science. I shall return to this problem in section four.

Even if PI had not been premature or allegedly transdisciplinary, the philosophical and scientific communities at large were not ready to appreciate its importance. There were strong programs of research, especially in various philosophies of language (logico-positivist, analytic, commonsensical, postmodernist, deconstructionist, hermeneutical, pragmatist, etc.), which attracted most of the intellectual and financial resources, kept a fairly rigid agenda which did not foster the evolution of alternative paradigms. Mainstream philosophy cannot help being conservative, not only because values and standards are usually less firm and clear in philosophy than in science, and hence more difficult to challenge, but also because, as we shall see better in section three, this is the context where a culturally dominant position is often achieved at the expense of innovative or unconventional approaches. As a result, thinkers like Church, Shannon, Engelbart, Simon, Turing, Von Neumann or Wiener were essentially left on the periphery of the traditional canon. Admittedly, the computational turn affected science much more rapidly. This explains why some philosophically-minded scientists were among the first to perceive the emergence of a new paradigm. But Sloman's "computer revolution" still had to wait until the 1980s to become a more widespread phenomenon across the various sciences and social contexts, thus creating the right environment for the emergence of the PI paradigm.

More than half a century after the construction of the first mainframes, society has now reached a stage in which issues concerning the creation, dynamics, management and utilisation of information and computational resources are vital. Nonetheless, advanced societies and western cultures had to undergo a revolution in digital communications before appreciating in full the radical novelty of the new

paradigm. The information society has been brought about by the fastest growing technology in history. No previous generation has ever been exposed to such an extraordinary acceleration of technological power over reality with the corresponding social changes and ethical responsibilities. Total pervasiveness, flexibility and high power have raised ICT to the status of the characteristic technology of our time, factually, rhetorically and even iconographically. The computer presents itself as a culturally defining technology and has become a symbol of the new millennium, playing a cultural role far more influential than the mills in the Middle Ages, mechanical clocks in the seventeenth century, or the steam engine in the age of the industrial revolution (Bolter 1984). ICS and ICT applications are nowadays the most strategic of all the factors governing science, the life of society and its future. The most developed post-industrial societies literally live by information, and ICS-ICT is what keeps them constantly oxygenated. And yet, all these profound and very significant transformations were barely in view two decades ago, when most philosophy departments would have considered topics in PI unsuitable areas of specialisation for a graduate student.

Too far ahead of its time, and dauntingly innovative for the majority of professional philosophers, PI wavered for some time between two alternatives. It created a number of interesting but limited research niches like philosophy of AI or computer ethics, often tearing itself away from its intellectual background. Or it was absorbed within other areas as a methodology, when PI was perceived as a computational or information-theoretic approach to otherwise traditional topics, in classic areas like epistemology, logic, ontology, philosophy of language, philosophy of science, or philosophy of mind. Both trends further contributed to the emergence of PI as an independent field of investigation.

2. The Historical Emergence of PI

Ideas, as it is said, are 'in the air'. The true explanation is presumably that, at a certain stage in the history of any subject, ideas become visible, though only to those with keen mental eyesight, that not even those with the sharpest vision could have perceived at an earlier stage (Dummett, 1993, 3).

Visionaries have a hard life. Recall Gide's image: if nobody else follows, one does not discover new lands but merely gets lost, at least in the eyes of those who stayed behind in Plato's cave. It has required a third computer-related revolution (the

networked computer, after the mainframe and the PC), a new generation of computer-literate students, teachers and researchers; a substantial change in the fabric of society, a radical transformation in the cultural and intellectual sensibility, and a widespread sense of crisis in philosophical circles of various orientations, for the new paradigm to emerge. By the late 1980s, PI had finally begun to be acknowledged as a fundamentally innovative area of philosophical research. Perhaps it is useful to recall a few dates. In 1982, *Time Magazine* named the computer “Man of the Year”. In 1985, the American Philosophical Association created the *Committee on Philosophy and Computers* (PAC). The “computer revolution” had affected philosophers as “professional knowledge-workers” even before attracting all their attention as interpreters. The charge of the APA Committee was, and still is, mainly practical. The Committee

collects and disseminates information on the use of computers in the profession, including their use in instruction, research, writing, and publication, and makes recommendations for appropriate actions of the Board or programs of the Association (from PAC web site).

Still in 1985, Terrell Ward Bynum, editor of *Metaphilosophy*, published a special issue of the journal entitled *Computers and Ethics* (Bynum 1985) that “quickly became the widest-selling issue in the journal’s history” (Bynum 2000, see also Bynum 1998). In 1986, the first conference sponsored by the *Computing and Philosophy* (CAP) association was held at Cleveland State University.

Its program was mostly devoted to technical issues in logic software. Over time, the annual CAP conferences expanded to cover all aspects of the convergence of computing and philosophy. In 1993, Carnegie Mellon became a host site. (from CAP web site).

It is clear that by the mid-1980s, the philosophical community was increasingly aware and appreciative of the importance of the topics investigated by PI, and of the value of its methodologies and theories (see for example Burkholder 1992, a collection of 16 essays by 28 authors presented at the first six CAP conferences; most of the papers are from the fourth). PI was no longer seen as weird, esoteric, transdisciplinary or philosophically irrelevant, or as a branch of applied IT. Concepts or processes like algorithm, automatic control, complexity, computation, distributed network, dynamic system, implementation, information, feedback or symbolic representation; phenomena like HCI (human-computer interaction), CMC (computer-mediated communication), computer crimes, electronic communities, or digital art; disciplines like AI or Information Theory; questions concerning the

nature of artificial agents, the definition of personal identity in a disembodied environment and the nature of virtual realities; models like those provided by Turing Machines, artificial neural networks and artificial life systems... these are just a few examples of a growing number of topics increasingly perceived as new, useful, of pressing interest and academically respectable. Informational and computational concepts, methods, techniques and theories had become powerful metaphors acting as “hermeneutic devices” through which to interpret the world. They had established a unified language that had become common currency in all academic subjects, including philosophy.

In 1998, exactly twenty years after the publication of Sloman’s *The Computer Revolution in Philosophy*, Terrell Ward Bynum and James H. Moor edited *The Digital Phoenix*, a collection of essays, this time significantly subtitled *How Computers are Changing Philosophy*. In the introduction, they acknowledged PI as 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. (Bynum and Moor 1998, p. 1).

At the short-sighted distance set by a textbook, philosophy often strikes the student as a discipline of endless diatribes and extraordinary claims, in a state of chronic crisis. *Sub specie aeternitatis*, the diatribes unfold in the forceful dynamics of ideas, claims acquire the necessary depth, the proper level of justification and their full significance, while the alleged crisis proves to be a fruitful and inevitable dialectic between innovation and scholasticism. This dialectic of reflection, highlighted by Bynum and Moor, has played a major role in establishing PI as a mature area of philosophical investigation. We have seen its historical side. Let us now see how it may be interpreted conceptually.

3. The Dialectic of Reflection and the Emergence of PI

In order to emerge and flourish, the mind needs to make sense of its environment by continuously investing data (constraining affordances, see chapter 5) with meaning. Mental life is thus the result of a successful reaction to a primary *horror vacui semantici*: meaningless (in the non-existentialist sense of “not-yet-meaningful”) chaos threatens to tear the Self asunder, to drown it in an alienating otherness perceived by the Self as nothingness. This primordial dread of annihilation urges the Self to go on filling any semantically empty space with whatever meaning the Self can muster, as successfully as inventiveness and the cluster of contextual constraints, affordances and the development of culture permit. This semanticisation of being, or reaction of the Self to the non-Self (to phrase it in Fichtean terms), consists in the inheritance and further elaboration, maintenance, and refinement of factual narratives (personal identity, ordinary experience, community ethos, family values, scientific theories, common-sense-constituting beliefs, etc.) that are logically and contextually (and hence sometimes fully) constrained and constantly challenged by the data that they need to accommodate and explain. Historically, the evolution of this process is ideally directed towards an ever-changing, richer and robust framing of the world. Schematically, it is the result of four conceptual thrusts:

1) a metasemanticisation of narratives. The result of any reaction to being solidifies into an external reality facing the new individual Self, who needs to appropriate narratives as well, now perceived as further constraining affordances that the Self is forced to semanticise. Reflection turns to reflection and recognises itself as part of the reality it needs to semanticise;

2) a de-limitation of culture. This is the process of externalisation and sharing of the conceptual narratives designed by the Self. The world of meaningful experience moves from being a private, infra-subjective and anthropocentric construction to being an increasingly inter-subjective and de-anthropocentrified reality. A community of speakers shares the precious semantic resources needed to make sense of the world by evolving and transmitting a language—with its conceptual and cultural implications—which a child learns as quickly as a shipwrecked person desperately grabs a floating plank. Narratives then become increasingly friendly because shared with other non-challenging Selves not far from one Self, rather than

reassuring because inherited from some unknown deity. As “producers” (producers and consumers) of specific narratives no longer bounded by space or time, members of a community constitute a group only apparently trans-physical, in fact functionally defined by the semantic space they wish and opt to inhabit. The phenomenon of globalisation is rather a phenomenon of erasure of old limits and creation of new ones, and hence a phenomenon of de-limitation of culture;

3) a de-physicalisation of nature. The physical world of shoes and cutlery, of stones and trees, of cars and rain, of the I as ID (the socially identifiable Self, with gender, job, driving license, marital status etc.) undergoes a process of virtualisation and distancing. Even the most essential tools, the most dramatic experiences or the most touching feelings, from war to love, from death to sex, can be framed within virtual mediation, and hence acquire an informational aura. Art, goods, entertainment, news and other Selves are placed and experienced behind a screen which is no longer an internal forum but a digital window. On the other side of this virtual frame, objects and individuals can become fully replaceable and often indistinguishable tokens of ideal types: a watch is really a swatch, a pen is a present only insofar as it is a branded object, a place is perceived as a holiday resort, a temple turns into a historical monument, someone is a police officer, and a friend may be just a written voice on the screen of a PC. Individual entities are used as disposable instantiations of universals. The here-and-now is transformed and expanded. By speedily multitasking, the individual Self can inhabit ever more *loci*, in ways that are perceived synchronically even by the Self, and thus swiftly weave different lives, which do not necessarily merge. Past, present and future are reshaped in discrete and variable intervals of current time. Projections and indiscernible repetitions of present events expand them into the future; future events are predicted and pre-experienced in anticipatory presents; while past events are registered and re-experienced in replaying presents. The nonhuman world of inimitable things and unrepeatably events is increasingly windowed and humanity window-shops in it;

4) a hypostatisation (embodiment) of the conceptual environment designed and inhabited by the mind. Narratives, including values, ideas, fashions, emotions and that intentionally privileged macro-narrative that is the I, can be shaped and reified into “semantic objects” or “information entities”. They now come closer to the

interacting Selves, quietly acquiring an ontological status comparable to that of ordinary things like clothes, cars and buildings.

By de-physicalising nature and embodying narratives, the physical and the cultural are re-aligned on the line of the virtual. In light of this dialectic, the information society can be seen as the most recent, although certainly not definitive, stage in a wider semantic process that makes the mental world increasingly part of, if not *the* environment in which more and more people tend to live. It brings history and culture, and hence time, to the fore as the result of human deeds, while pushing nature, as the non-human, and hence physical space, into the background.

In the course of its evolution, the process of semanticisation gradually leads to a temporal fixation of the constructive conceptualisation of reality into a world view, which then generates a conservative closure, scholasticism (for an enlightening discussion of contemporary scholasticism, see Rorty 1982, chaps. 2, 4 and esp. chap. 12).

Scholasticism, understood as an intellectual typology rather than a scholarly category, represents the inborn inertia of a conceptual system, when not its rampant resistance to innovation. It is *institutionalised philosophy* at its worst – a degeneration of what socio-linguists call, more broadly, the internal “discourse” (Gee 1998, esp. 52-53) of a community or group of philosophers. It manifests itself as a pedantic and often intolerant adherence to some discourse (teachings, methods, values, viewpoints, canons of authors, positions, theories or selections of problems etc.), set by a particular group (a philosopher, a school of thought, a movement, a trend, etc.), at the expense of alternatives, which are ignored or opposed. It fixes, as permanently and objectively as possible, a toolbox of philosophical concepts and vocabulary suitable for standardizing its discourse (its special *isms*) and the research agenda of the community. In this way, scholasticism favours the professionalisation of philosophy: scholastics are “lovers” who detest the idea of being *amateurs* and wish to become professional. They are suffixes. They call themselves “-ans” and place-before (pro-stituere) that ending the names of other philosophers, whether they are Aristotelians, Cartesians, Kantians, Nietzscheans, Wittgensteinians, Heideggerians or Fregeans. Followers, exegetes and imitators of some mythicized founding fathers, scholastics find in their hands more substantial answers than new interesting questions and thus gradually become involved with the application of

some doctrine to its own internal puzzles, readjusting, systematising and tidying up a once-dynamic area of research. Scholasticism is metatheoretically acritical and hence reassuring. Fundamental criticism and self-scrutiny are not part of the scholastic discourse, which, on the contrary, helps a community to maintain a strong sense of intellectual identity and a clear direction in the efficient planning and implementation of its research and teaching activities. It is also a closed context. Scholastics tend to interpret, criticise and defend only views of other identifiable members of the community, thus mutually reinforcing a sense of identity and purpose, instead of addressing directly new conceptual issues that may still lack an academically respectable pedigree and hence be more challenging. This is the road to anachronism. A progressively wider gap opens up between philosophers' problems and philosophical problems. Scholastic philosophers become busy with narrow and marginal *disputationes* of detail, while failing to interact with other disciplines, new discoveries, or contemporary problems that are of lively interest outside the specialised discourse. In the end, once scholasticism is closed in on itself, its main purpose becomes quite naturally the perpetuation of its own discourse, transforming itself into academic strategy.

Perhaps a metaphor can help to clarify the point. Conceptual areas are like mines. Some of them are so vast and rich that they will keep philosophers happily busy for generations. Others may seem exhausted until new and powerful methods or theories allow further and deeper explorations, or lead to the discovery of problems and ideas previously overlooked. Scholastic philosophers are like wretched workers digging a nearly exhausted but not yet abandoned mine. They belong to a late generation, technically trained to work only in the narrow field in which they happen to find themselves. They work hard to gain little, and the more they invest in their meagre explorations, the more they stubbornly bury themselves in their own mine, refusing to leave their place to explore new sites. Tragically, only time will tell whether the mine is truly exhausted. Scholasticism is a censure that can be applied only *post mortem*.

What has been said so far should not be confused with the naive question as to whether philosophy has lost, and hence should regain, contact with people (Adler 1979, Quine 1979). People may be curious about philosophy, but only a philosopher can fancy they might be deeply interested in it. It should also be distinguished from

questions of popularity. Scholasticism, if properly trivialised, can be pop, accessible and even trendy—after all, “trivial” should remind one of professional love.

Innovation is always possible, but scholasticism is historically inevitable. Any stage in the semanticisation of being is destined to be initially innovative if not disruptive, to establish itself as a specific dominant paradigm, and hence to become fixed and increasingly rigid, further reinforcing itself, until it finally acquires an intolerant stance towards alternative conceptual innovations, and so becomes incapable of dealing with the ever-changing intellectual environment that it helped to create and mould. In this sense, every intellectual movement generates the conditions of its own senescence and replacement.

Conceptual transformations should not be too radical, lest they become premature. We saw this at the beginning of section one. We have also seen that old paradigms are challenged and finally replaced by further, innovative reflection only when the latter is sufficiently robust to be acknowledged as a better and more viable alternative to the previous stage in the semanticisation of being. Here is how Moritz Schlick clarified this dialectic at the beginning of a paradigm shift:

Philosophy belongs to the centuries, not to the day. There is no uptodateness about it. For anyone who loves the subject, it is painful to hear talk of ‘modern’ or ‘non-modern’ philosophy. The so-called fashionable movements in philosophy—whether diffused in journalistic form among the general public, or taught in a scientific style at the universities—stand to the calm and powerful evolution of philosophy proper much as philosophy professors do to philosophers: the former are learned, the latter wise; the former write about philosophy and contend on the doctrinal battlefield, the latter philosophise. The fashionable philosophic movements have no worse enemy than true philosophy, and none that they fear more. When it rises in a new dawn and sheds its pitiless light, the adherents of every kind of ephemeral movement tremble and unite against it, crying out that philosophy is in danger, for they truly believe that the destruction of their own little system signifies the ruin of philosophy itself. (Schlick 1979, vol. II, 491)

Three types of forces therefore need to interact to compel a conceptual system to innovate. Scholasticism is the internal, negative force. It gradually fossilises thought, reinforcing its fundamental character of immobility and, by making a philosophical school increasingly rigid, less responsive to the world and more brittle, it weakens its capacity for reaction to scientific, cultural and historical inputs, divorces it from reality and thus prepares the ground for a solution of the crisis. Scholasticism, therefore, can indicate that philosophical research has reached a stage when it needs to address new topics and problems, adopt innovative methodologies, or develop alternative explanations. It does not, however, specify which direction the

innovation should take. Historically, this is the task of two other positive forces for innovation, external to any philosophical system: the substantial novelties in the environment of the conceptual system, occurring also as a result of the semantic work done by the old paradigm itself; and the appearance of an innovative paradigm, capable of dealing with them more successfully, and thus of disentangling the conceptual system from its stagnation. The rest of this section concentrates on the first positive force. The second one is discussed in section four.

In the past, philosophers had to take care of the whole chain of knowledge production, from raw data to scientific theories. Throughout its history, philosophy has progressively identified classes of empirical and logico-mathematical problems and outsourced their investigations to new disciplines. It has then returned to these disciplines and their findings for controls, clarifications, constraints, methods, tools and insights. However, *pace* Carnap (1935) and Reichenbach (1951), philosophy itself consists of conceptual investigations whose essential nature is neither empirical nor logico-mathematical. To mis-paraphrase Hume: “if we take in our hand any volume, let us ask: Does it contain any abstract reasoning concerning quantity or number? Does it contain any experimental reasoning concerning matter of fact and existence?” If the answer is yes, then search elsewhere, because that is science, not philosophy. Philosophy is not a conceptual aspirin, a super-science or the manicure of language. It is the last stage of reflection, where the semanticisation of being is pursued and kept open (Russell 1912, chap. 15). Philosophy’s creative and critical investigations identify, formulate, evaluate, clarify, interpret and explain conceptual problems that are intrinsically capable of different and possibly irreconcilable solutions, problems that are genuinely open to debate and honest disagreement, even in principle. These investigations are often entwined with empirical and logico-mathematical issues and so scientifically constrained but, in themselves, they are neither. They design and evaluate information environments and explanatory models, and thus 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 genuine, reasonable disagreement is always possible. It is also a dynamic space, for when its cultural environment changes, philosophy follows suit and evolves.

This normative space should not be confused with Sellars' famous "space of reasons":

in characterizing an episode or a state as that of knowing, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons of justifying and being able to justify what one says (Sellars 1963, 169).

Philosophy's normative space is a space of design, where rational and empirical affordances, constraints, requirements and standards of evaluation all play an essential role in the construction and evaluation of information and knowledge. It only partly overlaps with Sellars' space of reasons in that the latter includes more (e.g. mathematical deduction counts as justification, and in Sellars' space we find intrinsically decidable problems) and less, since in the space of information design we find issues connected with creativity and freedom not clearly included in Sellars' space (on Sellars' "space of reasons" see Floridi 1996, esp. chapter 4, and McDowell 1994, esp. the new introduction).

In Bynum's and Moor's felicitous metaphor, philosophy is indeed like a phoenix. It can flourish only by constantly re-engineering itself. A philosophy that is timeless instead of timely, rather than being an impossible *philosophia perennis*, which claims universal validity over past and future intellectual positions, is a stagnant philosophy, unable to contribute, keep track of, and interact with, the cultural evolution that philosophical reflection itself has helped to bring about, and hence to grow.

The more philosophy outsource various forms of knowledge, the more its pulling force has become external. This is the full sense in which Hegel's metaphor of the Owl of Minerva is to be interpreted. In the past, the external force has been represented by factors such as Christian theology, the discovery of other civilisations, the scientific revolution, the foundational crisis in mathematics and the rise of mathematical logic, evolutionary theory, and the theory of relativity, just to mention a few obvious examples. Nowadays, the pulling force of innovation is the complex world of information and communication phenomena, their corresponding sciences and technologies, together with the new environments, social life, existential and cultural issues that they have brought about. This is why PI can present itself as an innovative paradigm.

4. The Definition of PI

Once a new area of philosophical research is brought into being by the interaction between scholasticism and some external force, it evolves into a well-defined field, possibly interdisciplinary but still autonomous, only if:

- i) it is able to appropriate an explicit, clear and precise interpretation not of a scholastic *Fach* (Rorty 1982, chap. 2) but of the classic “*ti esti*”, thus presenting itself as a specific “philosophy of”;
- ii) the appropriated interpretation becomes a useful attractor for investigations in the new field;
- iii) the attractor proves sufficiently influential to withstand centrifugal forces that attempt to reduce the new field to other fields of research already well-established; and
- iv) the new field is rich enough to be organised in clear sub-fields and hence allow for specialisation.

Questions like “what is the nature of being?”, “what is the nature of knowledge?”, “what is the nature of right and wrong?”, “what is the nature of meaning?” are good examples of field-questions. They satisfy the previous conditions and guaranteed the stable existence of their corresponding disciplines. Other questions such as “what is the nature of the mind?”, “what is the nature of beauty and taste?”, or “what is the nature of a logically valid inference?” have been subject to fundamental re-interpretations, which have led to profound transformations in the definition of philosophy of mind, aesthetics and logic. Still other questions, like “what is the nature of complexity?”, “what is the nature of life?”, “what is the nature of signs?”, “what is the nature of control systems?” have turned out to be trans- rather than interdisciplinary. To the extent that the corresponding disciplines -- Complexity theory, Philosophy of Life, Semiotics and Cybernetics -- have failed to satisfy one or more of the previous conditions, they have struggled to establish themselves as academic, independent fields. The question is now whether PI itself satisfies (i)-(iv). A first step towards a positive answer requires a further clarification.

Philosophy appropriates the “*ti esti*” question essentially in two ways, *phenomenologically* (used here in its general meaning, to refer to the conceptual investigation of a related group of phenomena, and not to be confused with Husserl’s or Heidegger’s senses of phenomenology) or *metatheoretically*.

Philosophy of language and epistemology are two examples of “phenomenologies”. Their subjects are meaning and knowledge, not linguistic theories or cognitive sciences. The philosophy of physics and the philosophy of social sciences, on the other hand, are plain instances of “metatheories”. They investigate problems arising from organised systems of knowledge, which in their turn investigate natural or human phenomena. Some other philosophical branches, however, show only a *tension* towards the two poles, often combining phenomenological and metatheoretical interests. For example, this is the case with philosophy of mathematics and philosophy of logic. Like PI, their subjects are old, but they have acquired their salient features and become autonomous fields of investigation only very late in the history of thought. These philosophies show a tendency to work on specific classes of first-order phenomena, but they also examine these phenomena working their way through scientific theories concerning those phenomena. The tension pulls each specific branch of philosophy towards one or the other pole. Philosophy of logic, to rely on the previous example, is metatheoretically biased. It shows a constant tendency to concentrate primarily on conceptual issues arising from logic understood as a specific mathematical theory of formally valid inferences, whereas it pays much less attention to problems concerning logic as a natural phenomenon, what one may call, for want of a better description, rationality. *Vice versa*, PI, like philosophy of mathematics, is phenomenologically biased. It is primarily concerned with the domain of first-order phenomena represented by the world of information, computation and the information society. Nevertheless, it addresses its problems by starting from the vantage point represented by the methodologies and theories offered by ICS, and can incline towards a metatheoretical approach in so far as it is methodologically critical about its own sources.

We are now ready to discuss the following definition:

PI) philosophy of information (PI) is the philosophical field concerned with

- a) the critical investigation of the conceptual nature and basic principles of information, including its dynamics, utilisation and sciences, and
- b) the elaboration and application of information-theoretical and computational methodologies to philosophical problems.

The first half of the definition concerns philosophy of information as a new field. PI appropriates an explicit, clear and precise interpretation of the “*ti esti*” question, namely “What is the nature of information?”. This is the clearest hallmark of a new field. Of course, as with other field-questions, this only serves to demarcate an area of research, not to map its specific problems in detail (see Floridi 2001). As we shall see in chapter five, PI provides critical investigations that are not to be confused with a quantitative theory of data communication (information theory). On the whole, its task is to develop not a unified theory of information, but rather an integrated family of theories that analyse, evaluate and explain the various principles and concepts of information, their dynamics and utilisation. Special attention is paid to systemic issues arising from different contexts of application and the interconnections with other key concepts in philosophy, such as being, life, truth, knowledge, and meaning.

By “dynamics” of information the definition refers to:

PI.a.i) *the constitution and modelling of information environments*, including their systemic properties, forms of interaction, internal developments etc.;

PI.a.ii) *information life cycles*, i.e. the series of various stages in form and functional activity through which information can pass, from its initial occurrence to its final utilisation and possible disappearance. A typical life cycle includes the following phases: occurring (discovering, designing, authoring, etc.), processing and managing (collecting, validating, modifying, organising, indexing, classifying, filtering, updating, sorting, storing, networking, distributing, accessing, retrieving, transmitting, etc.) and using (monitoring, modelling, analysing, explaining, planning, forecasting, decision-making, instructing, educating, learning, etc.);

PI.a.iii) *computation*, both in the Turing-machine sense of *algorithmic processing*, and in the wider sense of *information processing*.

(PI.a.iii) introduces a crucial specification. Although a very old concept, information has finally acquired the nature of a primary phenomenon only thanks to the sciences and technologies of computation and ICT. Computation has therefore attracted much philosophical attention in recent years. Nevertheless, PI privileges “information” over “computation” as the pivotal topic of the new field because it analyses the latter as presupposing the former. PI treats “computation” as only one (although perhaps the most important) of the processes in which information can be involved. Thus,

the field should be interpreted as a philosophy of information rather than just of computation, in the same sense in which epistemology is the philosophy of knowledge, not just of perception. Thus, a shorter title for this volume could have been the *Blackwell Guide to PI*.

From an environmental perspective, PI is prescriptive about what may count as information, and how information should be adequately created, processed, managed and used (see chapter 6).

PI's phenomenological bias does not mean that it fails to provide critical feedback. On the contrary, methodological and theoretical choices in ICS are also profoundly influenced by the kind of explicit or implicit PI a researcher adopts. It is therefore essential to stress that PI critically evaluates, shapes and sharpens the conceptual, methodological and theoretical basis of ICS. In short, it also provides a *philosophy of ICS*, as has been plain since early work in philosophy of AI (see chapters 24-27).

An excessive concern with the metatheoretical aspects of PI may lead one to miss the important fact that it is perfectly legitimate to speak of PI even in authors who lived centuries before the information revolution. Hence, it will be extremely fruitful to develop a historical approach to trace PI's diachronic evolution. Technical and conceptual frameworks of ICS should not be anachronistically applied, but instead used to provide the conceptual method and privileged perspective to evaluate the reflections on the nature, dynamics and utilisation of information predating the digital revolution (e.g. Plato's *Phaedrus*, Descartes' *Meditations*, Nietzsche's *On the Use and Disadvantage of History for Life*, or Popper's conception of a third world). This is comparable to the development of other philosophical fields like philosophy of language, philosophy of biology, or philosophy of mathematics (for an interesting attempt to look at the history of philosophy from a computational perspective see Glymour 1992).

The second half of the definition (PI.b) indicates that PI is not only a new field, but introduces an innovative methodology as well. Research into the conceptual nature of information, its dynamics and utilisation is carried on from the vantage point represented by the methodologies and theories from ICS and ICT (chapter 27). This also affects the study of classic philosophical topics. Information-theoretic and computational methods, concepts, tools and techniques have already

been developed and applied in many philosophical areas, to extend our understanding of the cognitive and linguistic abilities of humans and animals, and the possibility of artificial forms of intelligence (chapters 10, 11, 17, 18); to analyse inferential and computational processes (chapters 19, 21, 22); to explain the organizational principles of life and agency (chapters 15, 16, 23); to devise new approaches to modelling physical and conceptual systems (chapters 12-14, 20); to formulate the methodology of scientific knowledge (chapters 24-26); to investigate ethical problems (chapter 6); aesthetic issues (chapter 9) and psychological, anthropological and social phenomena characterising the information society and human behaviour in digital environments (chapters 7-8). Indeed, the presence of these branches shows that PI satisfies criterion (iv). It provides a unified and cohesive theoretical framework that allows further specialisation.

PI possesses one of the most powerful conceptual vocabularies ever devised in philosophy. This is because we can rely on informational concepts whenever a complete understanding of some series of events is unavailable or unnecessary for providing an explanation. Virtually any issue can be rephrased in informational terms. This semantic power is a great advantage of the PI methodology. It is a sign that we are dealing with an influential paradigm, describable in terms of an informational philosophy. But it may also be a problem, because a metaphorically “pan-informational” approach can lead to a dangerous equivocation. Thinking that since anything can be described in (more or less metaphorically) informational terms, then everything has a genuinely informational nature. The risk is clear if one considers, for example, the difference between modelling the production chain that links authors, publishers and librarians as an information process, and representing digestion as if it were an information process. The equivocation obscures PI’s specificity as a philosophical field with its own subject. PI runs the risk of becoming synonymous with philosophy. A key that opens every lock only shows that there is something wrong with the locks. The best way of avoiding this loss of specificity is to concentrate on the first half of the definition. PI as a philosophical discipline is defined by what a problem is (or can be reduced to be) *about*, not by *how* it can be formulated. So although many philosophical issues may benefit greatly from an informational analysis, in PI information theory provides a literal foundation not just a metaphorical superstructure. PI presupposes that a problem or an explanation can

be legitimately and genuinely reduced to an informational problem or explanation. The criterion to test the soundness of the informational analysis of x is not to check whether x *can* be formulated in informational terms but to ask what would be like for x not to have an informational nature at all. With this criterion in mind, I have provided a sample of some interesting questions in Floridi 2001.

5. Conclusion: PI as *philosophia prima*

Philosophers have begun to address the new intellectual challenges arising from the world of information and the information society. PI attempts to expand the frontier of philosophical research, not by collating pre-existing topics, and thus reordering the philosophical scenario, but by forging new areas of philosophical inquiry and by providing innovative methodologies. Is the time ripe for the establishment of PI as a mature field? One may hope so. Our culture and society, the history of philosophy and the dynamic forces regulating the development of the philosophical system have been moving towards it. But then, what kind of PI can be expected to develop? An answer to this question presupposes a much clearer view of PI's position in the history of thought, a view probably obtainable only a posteriori. Here, it might be sketched by way of guesswork.

We have seen that philosophy grows by impoverishing itself. This is only an apparent paradox. The more complex the world and its scientific descriptions turn out to be, the more essential the level of the philosophical discourse understood as *philosophia prima* must become, ridding itself of unwarranted assumptions and misguided investigations that do not properly belong to the normative activity of conceptual modelling. The strength of the dialectic of reflection, and hence the crucial importance of historical awareness of it, lies in this transcendental regress in search of increasingly abstract and more streamlined conditions of possibility of the available narratives, in view not only of their explanation, but also of their modification and innovation. How has the regress developed? The *vulgata* suggests that the scientific revolution made seventeenth century philosophers redirect their attention from the nature of the knowable object to the epistemic relation between it and the knowing subject, and hence from metaphysics to epistemology. The subsequent growth of the information society and the appearance of the infosphere (the semantic environment which millions of people inhabit nowadays) led

contemporary philosophy to privilege critical reflection on the domain represented by the memory and languages of organised knowledge, the instruments whereby the infosphere is modelled and managed—thus moving from epistemology to philosophy of language and logic (Dummett 1993)—and then on the nature of its very fabric and essence, information itself. Information has thus arisen as a concept as fundamental and important as “being”, “knowledge”, “life”, “intelligence”, “meaning” or “good and evil”—all pivotal concepts with which it is interdependent—and so equally worthy of autonomous investigation. It is also a more basic concept, in terms of which the other can be expressed and interrelated, when not defined. In this sense, Evans was right:

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. That also distinguishes this work very sharply from traditional epistemology. (Dummett 1993, p. 186).

This is why PI can be introduced as a forthcoming *philosophia prima*, both in the Aristotelian sense of the primacy of its object, information, which PI claims to be a fundamental component in any environment, and in the Cartesian-Kantian sense of the primacy of its methodology and problems, since PI aspires to provide a most valuable, comprehensive approach to philosophical investigations.

PI, understood as a foundational philosophy of information modelling and design, can explain and guide the purposeful construction of our intellectual environment, and provide the systematic treatment of the conceptual foundations of contemporary society. It enables humanity to make sense of the world and construct it responsibly, reaching a new stage in the semanticisation of being. If what has been suggested here is correct, the current development of PI may be delayed but remains inevitable, and it will affect the overall way in which we address both new and old philosophical problems, bringing about a substantial innovation of the philosophical system. This will represent the information turn in philosophy. Clearly, PI promises to be one of the most exciting and fruitful areas of philosophical research of our time.

Acknowledgements

This chapter is a modified version of “What is the Philosophy of Information?”, an article published in T. W. Bynum and J. H. Moor (eds.), *CyberPhilosophy: The Intersection of Philosophy and Computing*, special issue of *Metaphilosophy*, volume 33, issues 1/2, January 2002. I am grateful to the publisher for permission to reproduce the text here.

References

- Adler, M. 1979, “Has Philosophy Lost Contact With People?” *Long Island Newsday*, November 18.
- Anderson, A. R. (ed.), 1964 *Minds and Machines, Contemporary Perspectives in Philosophy Series* (Englewood Cliffs: Prentice-Hall).
- Boden, M. A. 1984, “Methodological Links between AI and Other Disciplines” in *The Study of Information: Interdisciplinary Messages*, ed. by F. Machlup and V. Mansfield (New York, John Wiley and Sons), rep. in Burkholder 1992.
- Boden, M. A. (ed.), 1990 *The Philosophy of Artificial Intelligence, Oxford Readings in Philosophy* (Oxford: Oxford University Press).
- Bolter J. D. 1984, *Turing’s Man. Western Culture in the Computer Age* (Chapel Hill: The University of North Carolina Press).
- Burkholder, L. (ed.) 1992, *Philosophy and the Computer* (Boulder, San Francisco, Oxford: Westview Press).
- Bynum, T. W. (ed.), 1985 *Computers and Ethics* (Oxford: Blackwell, published as the October 1985 issue of *Metaphilosophy*).
- Bynum, T. W. 1998 “Global Information Ethics and the Information Revolution” in Bynum and Moor 1998, 274-289.
- Bynum, T. W. and Moor, J. H. (eds.), 1998 *The Digital Phoenix: How Computers are Changing Philosophy*, special issue of *Metaphilosophy* also available as a book (New York - Oxford: Blackwell).
- Bynum, T. W., 2000 “A Very Short History of Computer Ethics”, *APA Newsletters on Philosophy and Computers*, Spring, 99.2.
- Bynum, T. W. and Moor, J. H. (eds.), 2002 *Cyberphilosophy: The interasection of philosophy and computing* special issue of *Metaphilosophy* also available as a book (New York - Oxford: Blackwell).

- CAP, web site of the Computing and Philosophy annual conference series, <http://www.lcl.cmu.edu/caae/cap/CAPpage.html>
- Carnap, R. 1935, *Philosophy and Logical Syntax*, Chap. “The Rejection of Metaphysics”.
- Dummett, M. 1993, *Origins of Analytical Philosophy* (London: Duckworth).
- Floridi L. 1996, *Scepticism and the Foundation of Epistemology—A Study in the Metalogical Fallacies* (Leiden: Brill).
- Floridi, L. 2001, “Open Problems in the Philosophy of Information”, *The Herbert A. Simon Lecture on Computing and Philosophy*, Carnegie Mellon University, 10 August 2001, preprint available at <http://www.wolfson.ox.ac.uk/~floridi/papers.htm>
- Gee, J. P. 1998, “What is Literacy?”, in V. Zamel and R. Spack (eds.), *Negotiating Academic Literacies: Teaching and Learning Across Languages and Cultures* (Mahwah, NJ: Erlbaum), pp. 51-59.
- Glymour, C. N. 1997, *Thinking Things Through: An Introduction to Philosophical Issues and Achievements* (Cambridge, Mass.: MIT Press).
- Haugeland, J. (ed.), 1981 *Mind Design: Philosophy, Psychology, Artificial Intelligence* (Montgomery, Vt.: Bradford Books).
- Haugeland, J. (ed.), 1997 *Mind Design II: Philosophy, Psychology, Artificial Intelligence* (Cambridge, Mass.: MIT Press).
- McCarthy J. and Hayes P. J. 1969, “Some Philosophical Problems from the Standpoint of Artificial Intelligence”, *Machine Intelligence*, 4, 463-502.
- McCarthy J. 1995, “What has AI in Common with Philosophy?”, *Proceedings of the 14th International Joint Conference on AI*, Montreal, August 1995, <http://www-formal.stanford.edu/jmc/aiphil.html>
- McDowell J. 1994, *Mind and World* (Cambridge Ma: Harvard University Press).
- PAC, web site of the American Philosophical Association Committee on Philosophy and Computers, <http://www.apa.udel.edu/apa/governance/committees/computers/>
- Pagels, H. 1988, *The Dreams of Reason: The Computer and the Rise of the Sciences of Complexity* (New York: Simon and Schuster).
- Pylyshyn Z. W. (ed.) 1970, *Perspectives on the Computer Revolution* (Englewood Cliffs, NJ: Prentice-Hall).

- Quine, W. V. O. 1979, "Has Philosophy Lost Contact with People?" *Long Island Newsday*, November 18. The article was modified by the editor. The original version appears as essay n. 23 in *Theories and Things* (Cambridge, Mass.: Harvard University Press, 1981).
- Reichenbach, H. 1951, *The Rise of Scientific Philosophy* (Berkeley: University of California Press).
- Ringle, M. (ed.) 1979, *Philosophical Perspectives in Artificial Intelligence* (Atlantic Highlands N.J., Humanities Press).
- Rorty, R. 1982, *Consequences of Pragmatism* (Brighton: The Harvester Press).
- Russell, B. 1912, *The Problems of Philosophy* (Oxford: Oxford University Press).
- Schlick, M. 1979, "The Vienna School and Traditional Philosophy", Eng. tr. by P. Heath in Moritz Schlick, *Philosophical Papers*, 2 vols. (Dordrecht: Reidel, orig. 1937).
- Sellars, W. 1963, *Science, Perception and Reality* (London and New York: New York Humanities Press).
- Simon H. A. 1962, "The Computer as a Laboratory for Epistemology", first draft, revised and published in Burkholder 1992, pp. 3-23.
- Simon H. A. 1996, *The Sciences of the Artificial* 3rd ed. (Cambridge, Mass.: MIT Press).
- Sloman A. 1978, *The Computer Revolution in Philosophy* (Atlantic Highlands: Humanities Press).
- Sloman A. 1995, "A Philosophical Encounter - An Interactive Presentation of Some of the Key Philosophical Problems in AI and AI Problems in Philosophy", *Proceedings of the 14th International Joint Conference on AI*, Montreal, August 1995, http://www.cs.bham.ac.uk/~axs/cog_affect/ijcai95.text
- Stent, G. 1972, "Prematurity and Uniqueness in Scientific Discovery", *Scientific American*, December, 84-93.
- Torrance S. B. 1984, *The Mind and The Machine: Philosophical Aspects of Artificial Intelligence* (Chichester West Sussex - New York, Ellis Horwood Halsted Press).