INFORMATION ETHICS GROUP

Oxford University and University of Bari

How To Do Philosophy Informationally

by

Gian Maria Greco, Gianluca Paronitti, Matteo Turilli, and Luciano Floridi

gianmaria.greco@ateneo.unile.it

IEG – RESEARCH REPORT 04.09.04

http://web.comlab.ox.ac.uk/oucl/research/areas/ieg

Abstract

In this paper we introduce three methods to approach philosophical problems informationally: Minimalism, the Method of Abstraction and Constructionism. Minimalism considers the specifications of the starting problems and systems that are tractable for a philosophical analysis. The Method of Abstraction describes the process of making explicit the level of abstraction at which a system is observed and investigated. Constructionism provides a series of principles that the investigation of the problem must fulfil once it has been fully characterised by the previous two methods. For each method, we also provide an application: the problem of visual perception, functionalism, and the Turing Test, respectively.

Keywords

Constructionism, Informational Methodology, Level of Abstraction, Minimalism, Philosophy of Information.

1. Introduction

The Philosophy of Information is a new area of research that has recently been developed at the intersection of computer science and philosophy [1]. It concerns (a) the critical investigation of the conceptual nature and basic principles of information, including its dynamics (especially computation), utilization (especially computer ethics) and sciences; and (b) the elaboration and application of computational and information-theoretic methodologies to philosophical problems. Past work by members of our group has concentrated on (a), and in this paper we wish to explore (b). In a nutshell, we wish to ask not what philosophy can do for computer science, but rather what the latter can do for the former.

The applications of computational methods to philosophical issues may be approached in three ways [3]:

1) Conceptual experiments *in silico*, or the externalization of the mental theatre. As Patrick Grim has remarked "since the eighties, philosophers too have begun to apply computational modeling to questions in logic, epistemology, philosophy of science, philosophy of mind, philosophy of language, philosophy of biology, ethics, and social and political philosophy. […] A number of authors portray computer experimentation in general as a technological extension of an ancient tradition of thought experiment" [10].
2) *Pancomputationalism* or the fallacy of a powerful metaphor. Computational and informational concepts are so powerful that, given the right Level of Abstraction (see below), anything can be presented as a computational 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 computationally - heating, flying and knitting. Pancomputationalists (e.g. Chalmers) have the hard task of providing a credible answer to the question: what would it mean for the system under investigation not to be an informational system (i.e., a computational system, if computation = information processing)? Pancomputationalism does not seem vulnerable to a refutation, in the form of a possible counterexample in a world nomically identical to the one to which pancomputationalism is applied.
3) *Regulae ad directionem ingenii* or the Cartesian-Kantian approach. Are there specific methods in computer science that can help us to approach philosophical problems computationally? In the following pages we have tried to answer this question in the positive, by highlighting three main methods: *Minimalism*, the *Method of Abstraction* and *Constructionism*. Each method is discussed in a separate section.

2. Minimalism

Philosophical questions pose multi-faceted problems. Following Descartes, one can decompose a problem space with a divide-and-conquer approach. The outcome is a set of more approachable sub-

problems interconnected in a sort of Quinean web of dependencies. Often, the starting problem in answering a philosophical question presupposes other open problems. The strength of the answer depends on the strength of the corresponding assumptions. A minimalist starting problem relies as little as possible on other open problems. Consequently, a more robust answer to the philosophical question is more likely.

Philosophers improve the tractability of a problem space by choosing discrete systems with which it may be studied. Minimalism outlines three criteria to orientate this choice: *controllability*, *implementability* and *predictability*. A system is controllable when its structure can be modified purposefully. Given this flexibility, the system can be used as a case study to test different solutions for the problem space. The second minimalist criterion states that systems must be implemented physically or by simulation. The system becomes a white box – the opposite of a black box. Metaphorically, the maker of the system is like a Platonic "demiurge". She knows the components of the system and its state transitions rules and can use the system as a laboratory to test specific constraints on the problem space. The third criterion follows from the previous two: the behaviour of the system must be predictable. The demiurge can predict the behaviour of the system in the sense that she can infer the correct consequences from her explanations of the system. The system outcomes become then the benchmarks of the tested solutions.

The elements for the definition of minimalism are now available.

1. Minimalism is relational. Problems and systems are never absolutely minimalist, but always connected with the problem space posed by the philosophical question.
2. Minimalism provides a way to choose critically the starting problem for the analysis of a problem space. A minimalist starting problem relies as little as possible on other open problems and guarantees the strength of the next step in the forward process of answering the philosophical question.
3. According to a minimalist approach, the tractability of a philosophical problem is a function of the three criteria outlined previously. They allow the use of dynamic systems to test possible solutions and to derive properties of the problem space.
4. Finally, Minimalism is not a way to privilege simple or elementary problems. A minimalist problem may be difficult or complex. Minimalism is a matter of inferential relations between a problem and its space, as it seeks to guarantee that the chosen problem does not presuppose other open problems.

Minimalism is an economic method that may be confused with Ockham's razor. The two methods are compatible, but while Ockham's Razor avoids inconsistencies and ambiguities by eliminating redundant explicative or ontological elements in a theory, Minimalism provides a set of criteria for choosing problems and systems relative to a given specific question. Moreover, Ockham's

principle of parsimony is absolute and is applied to any theoretical element while Minimalism's main maxims of strength and tractability are always relative to a given problem space.

Let us now consider a practical example of Minimalism applied to the philosophy of perception.

1. The identification of the question: "What is visual perception?". This question poses a wide problem space, approached with different methods.
2. The Cartesian decomposition of the problem followed by a Quinean construction of the problem space. Well-known sub-problems of this problem space are the nature of internal representations, the role of mind in perception, vision as computation etc.
3. The identification of the starting problem. The standard representational interpretation of perception is rich in assumptions about open problems. Perception is based, for example, on the presumed existence of internal representations. The sensorimotor approaches to visual perception are less demanding. Perception is chained to action while information is externalised. James Gibson, one of the main advocates of the sensorimotor hypothesis, cannot explain the nature of perceptual errors. This problem does not rely on other open problems and therefore can be assumed as a minimalist starting problem: the "Gibson problem".
4. The selection of the system to be used to study the starting problem. This system has to be consistent with the requirements of Gibson's sensorimotor theory and with the criteria for Minimalism. The subsumption architecture proposed by Rodney Brooks fits these requirements. The architecture of Brooks' robots is reactive, parallel and decentralised. Perception and action are directly connected without any explicit internal representation or centralised inferential engines. Moreover, subsumption architectural behaviour is fully specified by the topological structure of its layers composed of single behavioural units. Its designer has full control and predictability power over the system she has built. The Gibson problem can then be studied by means of Brooks' mobots.
5. The solution of the problem. In the sensorimotor approaches to vision, seeing is something done by agents in their environments. The definition of perceptual errors must be shifted from a representational interpretation – errors are wrong computations made over internal representations – to an action-based interpretation – errors are unsuccessful actions made by agents in their environment. If the mobot's sensorimotor features enable it to move randomly in its environment then perception is successful. If not, then its perception is erroneous. The mobot that bumps against a window lacks the right features or the right kind of sensorimotor capabilities relative to a given specific environment and its task of moving around randomly.

3. The Method of Abstraction

The process of making explicit the Level of Abstraction at which a system is considered is called Method of Abstraction. This method applies both to conceptual and phenomenological systems and its

basic element is the concept of Level of Abstraction (LoA). The metaphor of interface in a computer system is helpful to illustrate what a LoA is. We all know that users seldom think about the fact that actually they use a variety of interfaces between them and the real electro-Boolean processes that carry out the required operations. An interface may be described as an intra-system, which transforms the outputs of one system *A* into the inputs of a system *B* and vice versa, producing a change in data types. LoAs are equivalent to interfaces because: (a) they are network of observables; (b) the observables are related by behaviours that moderate the LoA and can be expressed in terms of transition rules (c) they are conceptually positioned between data and the information spaces of the agents; (d) they are the place where (diverse) independent systems meet, act on or communicate with each other.

LoAs can be connected together to form broader structure of abstraction, from hierarchy of abstractions to nets of abstraction. One of the possible relation between LoAs is the one of simulation. A simulation relation is the relation between the observables of a simulator system and a simulated one. This relation must occur between pairs of observables in order to guarantee a certain congruence not only for the current state of the two systems but also for their evolution. In the simulation relation, the epistemic agent is coupling the state evolution of two systems by observing these two systems at different LoAs. This means that an epistemic agent tries to construct a sort of equivalence relation between the two systems, seeking to understand at what LoA those systems could be considered congruent. Let us now apply the Method of Abstraction to the case of functionalism.

Functionalism argues that a physical or abstract entity is identified by its causal or operational role. From this viewpoint, a system is not evaluated by its structures and their interactions, but rather by the functions it shows. If the "matter" constituting a system is irrelevant in order to identify it, then the same functional organization can be realized by different systems, which are usually called realizations. This is the so-called multirealizability thesis. Some philosophers try to rule out multirealizability from the functionalist approach. They argue that multirealizability could lead to a weakening of a neuro-scientific approach in the explanation of human behaviour. Why bother with actual neural structures if one can execute an algorithm to instantiate the same behaviours shown by these neural structures? Naturally, a computational approach is more suitable for processing those algorithms.

However, multirealizability cannot be detached from functionalism. Without multirealizability functionalism would become inexplicable. This is clear if we consider the mathematical concept of function. A function is usually expressed by an operation on one or more variables. The general-known scheme is $f(x) = y$, but this simply means that the variables in the equation could be realized by an infinite class of numbers or by points over the Cartesian plane or by means of a Turing machine or by set theory. Without all these instantiations, how could we explain the function $f(x) = y$? So let us assume that functionalism entails multirealizability. In the classic account of functionalism we have: the relata (the *functional organization* and the *realizations*) and the relations (the *realization relation* between the functional organization and the realizations and the *simulation relation* between the

various realization). We want to show that realization and simulation are equivalent. One can say that given that an epistemic agent can observe any functional organization at a specific LoA and the realization of that functional organization at another LoA, then the realization relation between the two LoAs is characterized by: (1) the process of codification of the inputs of the functional organization LoA into the inputs of the various realizations LoAs, and (2) the de-codification of the outputs of the second into the outputs of the first. Basically, simulation relation and realization relation are equivalent because they are relations which describe the same processes.

The argument is then that (a) multirealizability and functionalism are coupled concepts and (b) a simulation relation is equivalent to a realization relation. It follows that (c) a common functional organization does not exist at a higher LoA than its realizations. The functional organization is the Net of Abstraction constructed by the epistemic agents with the simulation relation between the various realizations conceived at different LoAs. This means that a functional organization is the relational structure produced by various realizations and by the simulation relation that connects those realizations. Perhaps a metaphor may help here. What the argument shows is that a carpenter who is making a piece of furniture by following a blueprint is not handling a functional organization (the blueprint) and a realization (the piece of furniture), but two realizations at different LoAs, which are related in a simulation relation specified by his work.

The new interpretation of functionalism just provided leads us to reconsider functionalistic explanations within the philosophy of AI and the philosophy of mind by introducing a new player, namely the simulation relation. This kind of explanation, given our interpretation of functionalism, should be configured as a specification of simulations between the LoAs at which the realizations are disposed by the epistemic agent.

4. Constructionism

A black box is a system whose internal structure, rules and composition remain undisclosed. A white box is a system about which one knows everything, because one has constructed it. This perspective lays in the wake of the so-called maker's knowledge tradition, according to which: "(a) One can only know what one makes. So, (b) one cannot know the genuine nature of reality in itself". Philosophers who stress (b) argue that, since any attempt to reach an understanding of the world will inevitably fail, it is better to concentrate on those sciences whose subject is created by us, such as politics and social sciences. Philosophers who stress (a) argue that it is possible to improve our knowledge of reality through the improvement of our knowledge of the techniques by which reality is investigated. This tradition finds its champion in Francis Bacon's Philosophy of Technology. With Bacon, Technology becomes the main subject of philosophical enquiry, because it is both a human product and the means through which the world is investigated. Constructionism explicitly refers to the maker's knowledge tradition. Its method consists in the following five principles:

1. The *Principle of Knowledge*: only what is constructible is knowable. Anything that can not be constructed could be subject, at most, to a working hypothesis.
2. The *Principle of Constructability*: working hypotheses are investigated by (theoretical or practical) simulations based on them.
3. The *Principle of Controllability*: simulations must be controllable.
4. The *Principle of Confirmation*: any confirmation or refutation of the hypothesis concerns the simulation, not the simulated.
5. The *Principle of Economy*: the less conceptual resources are used, the better it is. In any case, the resources used must be less than the results accomplished.

Newell and Simon outlined clearly the constructionist approach in computer science by saying that: "neither machines nor programs are black boxes; they are artefacts that have been designed, both hardware and software, and we can open them up and look inside" [11]. Constructionism suggests that, given a theory, one implements and tests it in a system. Because one constructs the system, one can also control it. Consider behaviour-based robotics. One observes an ant, make a hypothesis on its internal structures to explain its behaviours, then one builds a system to test that hypothesis. The resulting system is controllable in that it is modifiable, compositional and predictable. This means that, as far as the constructed system is concerned, one can change its internal structures and rules; the system can be implemented by adding or removing new parts; and since one knows the rules of the system, one can know its behaviour. Suppose that the mobot we have constructed behaves like an ant. The Principle of Confirmation prevents us from generalizing the working hypotheses, as if the simulation were the real cause (or internal structure) of the simulated. From this, the *sub-Principle of Context-dependency* derives: isomorphism between the simulated and simulation is only local, not global. The mobot accounts for the behaviour of the ant only under the constraints specified by the simulation. If the constraints change, so does the evaluation of the hypotheses.

All this is in plain opposition to mimetic theories of knowledge. These assume that reality is the target of our knowledge, which we acquire through some kind of mimetic mechanism: ideas, mental images, corresponding pictures and so forth. Adopting a constructionist point of view means rejecting mimetic theories such as Plato's, Descartes' or Locke's. The Principle of Economy refers to the "careful management of resources". On the one hand, in defining knowledge processes, mimetic theories use a large amount of resources. Assuming that there is a reality and that it works in such and such way means making a heavy ontological commitment. On the other hand, Constructionism does not state anything about reality in itself. A more modest commitment makes errors less likely.

The Turing Test (TT) is a perfect example to show how the methodology and, more specifically, the constructionist method work, for it respects the minimalist criterion, uses the LoAs and is constructionist.

Turing refuses even to try to provide an answer to the question "can a machine think?". He considers it a problem "too meaningless to deserve discussion" [12], because it involves such vague concepts as 'machine' and 'thinking'. Turing suggests replacing it with the Imitation Game, which is exactly more manageable and less demanding from the minimalist point of view. By so doing, Turing specifies a LoA and asks a new question, which may be summed up thus: "can we consider that a digital computer is thinking at this Level of Abstraction?". The rules of the game define the conditions of observability. If we observe the behaviour under those conditions, we can accept an operational definition of thinking machine for that LoA. By changing the rules of the game one changes the LoA so the answer will change too.

Note how TT respects the constructionist principles:

1. By satisfying minimalism, Turing also respects the Principle of Knowledge.
2. Turing makes a hypothesis based on the common assumption that conversation skills require intelligence, and then he devises a system to evaluate whether a machine is intelligent comparatively.
3. The system is controllable. It is known how it works and it can be modified.
4. Whether a machine passes the test implies only that the machine can, or can not, be considered intelligent at that LoA.
5. Finally, in tackling the problem of Artificial Intelligence, Turing refuses to consider those ways requiring a large amount of conceptual resources. This is, for instance, why he refuses to deal with any psychological assumption about intelligence.

5. Conclusion

In this paper, we have introduced three methods and shown how they can be successfully and fruitfully imported from computer science into philosophy, in order to model, analyse and solve conceptual problems. We have demonstrated their principal features and main advantages. The methods clarify implicit assumptions, facilitate comparisons, enhance rigour and promote the resolution of possible conceptual confusions. Of course, the adoption of the methods raises important further questions. By way of conclusion, we wish to call the reader's attention to only three of them that seem to us particularly pressing:

1. What is the logic of problem spaces?
2. What are the logical relations between LoAs?
3. How can constructionism avoid solipsism?

We have not attempted to answer these questions, which we hope to address in a future work.

Some applications of the methods discussed in this paper have already been provided in the philosophy of mind [6], in computer ethics [2] [9], in epistemology [4] [8], in the philosophy of science [7] and in the philosophy of information [5]. In each case, the methods have been shown to provide a flexible and fruitful approach. But it is obvious that much more work lies ahead[1].

Bibliography

1. Floridi, L.: What Is the Philosophy of Information?. Metaphilosophy. 1-2 (2002) 123-45
2. Floridi, L.: On the Intrinsic Value of Information Objects and the Infosphere. Ethics and Information Technology. 4 (2003) 287-304
3. Floridi, L.: Two Approaches to the Philosophy of Information. Minds and Machines. 13 (2003) 459-69
4. Floridi, L.: On the Logical Unsolvability of the Gettier Problem. Synthese. (2004)
5. Floridi, L.: Open Problems in the Philosophy of Information. Metaphilosophy. 4 (2004) 554-82
6. Floridi, L. Consciousness, Agents and the Knowledge Game. Forthcoming
7. Floridi, L.: The Informational Approach to Structural Realism. Forthcoming-b
8. Floridi, L.: Presence: From Epistemic Failure to Successful Observability. Presence: Teleoperators and Virtual Environments. Forthcoming-c
9. Floridi, L., Sanders, J. W.: On the Morality of Artificial Agents. Minds and Machines. (2004)
10. Grim, P.: Computational Modeling as a Philosophical Methodology. In Floridi, L. (ed.): The Blackwell Guide to the Philosophy of Computing and Information. Blackwell, Oxford (2004)
11. Newell, A., Simon, H. A.: Computer Science as Empirical Enquiry: Symbols and Search. Communications of the ACM. 3 (1976) 113-126
12. Turing, A. M.: Computing Machinery and Intelligence. Mind. 49 (1950) 433-460

[1] For Italian legal requirements, Gianluca Paronitti must be considered the author of section 3, Matteo Turilli of section 2, Luciano Floridi of sections 1 and 5, Gian Maria Greco of section 4 and the first author of the whole paper. As usual, Jeff Sanders' input was fundamental in shaping our ides.