The Method of Levels of Abstraction

Luciano Floridi[1, 2]

[1]Research Chair in Philosophy of Information and GPI, University of Hertfordshire; [2]St Cross College and IEG, University of Oxford.

Address for correspondence: School of Humanities, University of Hertfordshire, de Havilland Campus, Hatfield, Hertfordshire AL10 9AB, UK; l.floridi@herts.ac.uk

Abstract

The use of “levels of abstraction” in philosophical analysis (*levelism*) has recently come under attack. In this paper, I argue that a refined version of *epistemological levelism* should be retained as a fundamental method, called *the method of levels of abstraction*. After a brief introduction, in section two the nature and applicability of the epistemological method of levels of abstraction is clarified. In section three, the philosophical fruitfulness of the new method is shown by using Kant’s classic discussion of the “antinomies of pure reason” as an example. In section four, the method is further specified and supported by distinguishing it from three other forms of “levelism”: (i) levels of organisation; (ii) levels of explanation and (iii) conceptual schemes. In that context, the problems of relativism and antirealism are also briefly addressed. The conclusion discusses some of the work that lies ahead, two potential limitations of the method and some results that have already been obtained by applying the method to some long-standing philosophical problems.

Keywords

Abstraction, level of abstraction, gradient of abstraction, levelism, observable, stance.

1. Introduction

Reality can be studied at different levels, so forms of "levelism" have often been advocated in the past.[1] In the seventies, levelism nicely dovetailed with the computational turn and became a standard approach both in science and in philosophy. Dennett [1971], Mesarovic et al. [1970], Simon [1969] (see now Simon [1996]) and Wimsatt [1976] were among the earliest advocates. The trend reached its acme at the beginning of the eighties, with the work of Marr [1982] and Newell [1982]. Since then, levelism has enjoyed great popularity[2] and even textbook status (Foster [1992]). However, after decades of useful service, levelism seems to have come under increasing criticism.

Consider the following varieties of levelism currently available in the philosophical literature:

1) epistemological, e.g., levels of observation or interpretation of a system (see section four);
2) ontological, e.g., levels (or rather layers) of organization, complexity, or causal interaction etc. of a system;[3]
3) methodological, e.g., levels of interdependence or reducibility among theories about a system; and
4) an amalgamation of (1)-(3), e.g., as in Oppenheim and Putnam [1958].

The current debate on multirealizability in the philosophy of Artificial Intelligence (AI) and cognitive science has made (3) controversial, as Block [1997] has shown. And two recent articles by Heil [2003] and Schaffer [2003] have seriously and convincingly questioned the plausibility of (2). Since criticisms of (2) and (3) end up undermining (4), rumours are that levelism should probably be decommissioned.

In this paper, I agree with Heil and Schaffer that *ontological levelism* is probably

[1] See for example Brown [1916]. Of course the theory of ontological levels and the "chain of being" goes as far back as Plotin and forms the basis of at least one version of the ontological argument.

[2] The list includes Arbib [1989], Bechtel and Richardson [1993], Egyed and Medvidovic [2000], Gell-Mann [1994], Kelso [1995], Pylyshyn [1984], Salthe [1985].

[3] Poli [2001] provides a reconstruction of ontological levelism; more recently, Craver [2004] has analysed ontological levelism, especially in biology and cognitive science, see also Craver [forthcoming].

untenable. However, I shall also argue that *epistemological levelism* should be retained as a fundamental and indispensable method of conceptual analysis, if in a suitably refined version. Fleshing out and defending epistemological levelism is the main task of this paper, where I shall outline a theory of levels of abstraction. This is achieved in two stages. First, I shall clarify the nature and applicability of what I shall call *the method of* (*levels of*) *abstraction*. Second, I shall distinguish this method from other level-based approaches, which may not, and indeed need not, be rescued. Here is a more detailed overview of the paper.

In section two, I provide a definition of the basic concepts fundamental to the method. Although the definitions require some rigour, all the main concepts are introduced without assuming any previous knowledge. The definitions are illustrated by several intuitive examples, which are designed to familiarise the reader with the method.

In section three, I show how the method of abstraction may be fruitfully applied to philosophical topics by using Kant's discussion of the "antinomies of pure reason".

In section four, I further specify and support the method of abstraction by distinguishing it from three forms of "levelism": (i) ontological levels of organisation; (ii) methodological levels of explanation and (iii) conceptual schemes. In that context, I also briefly address the problems of relativism and antirealism.

In the conclusion, I indicate some of the work that lies ahead, two potential limitations of the method and some interesting results that have already been obtained by applying the method to some long-standing philosophical problems in different areas.

Before starting, one last bit of information and an acknowledgement of my intellectual debts are in order. The bit of information concerns a second paper, on the same topic, which contains several specific applications of the method of abstraction illustrating its fruitfulness. Despite some redundancy between the two papers, the reader interested in the topic might be curious to check Floridi [forthcoming-b]. As for the debt, many of the ideas presented here were developed in collaboration with Jeff Sanders (Floridi and Sanders [2004a]). Although levelism has been common currency in philosophy and in science since antiquity, only more recently has the concept of *simulation* been used in computer science to relate levels of abstraction to satisfy the requirement that systems constructed in levels (in order to tame their complexity) function correctly (see for example De Roever and Engelhardt [1998], Hoare and He [1998]). The definition of *Gradient of Abstraction* (GoA, see section 2.6) has been inspired by this approach. Indeed, I take as a definition the property established by simulations, namely the conformity of behaviour between levels of abstraction (more on this later).

2. Some Definitions and Preliminary Examples

In this section, I introduce six key concepts – namely, “typed variable”, “observable”, “level of abstraction”, “behaviour”, “moderated level of abstraction” and “gradient of abstraction” – some simple examples to illustrate their use, and then the “method of abstraction” based on them.

2.1. Typed Variable

As is well known, a variable is a symbol that acts as a place-holder for an unknown or changeable referent. In this article, a “typed variable” is a variable qualified to hold only a declared kind of data.

> Definition: A *typed variable* is a uniquely-named conceptual entity (the *variable*) and a set, called its *type*, consisting of all the values that the entity may take. Two typed variables are regarded as *equal* if and only if their variables have the same name and their types are equal as sets. A variable that cannot be assigned well-defined values is said to constitute an *ill-typed variable* (see the example in section 2.3).

When required, I shall write x:X to mean that x is a variable of type X. Positing a typed variable means taking an important decision about how its component variable is to be conceived. This point may be better appreciated after the next definition.

2.2. Observable

The notion of an "observable" is common in science, occurring whenever a (theoretical) model is constructed. Although the way in which the features of the model correspond to the system being modelled is usually left implicit in the process of modelling, it is important here to make that correspondence explicit. I shall follow the standard practice of using the word "system" to refer to the object of study. This may indeed be what would normally be described as a system in science or engineering, but it may also be a domain of discourse, of analysis, or of conceptual speculation: a purely semantic system, as it were.

> Definition: An *observable* is an interpreted typed variable, that is, a typed variable together with a statement of what feature of the system under consideration it represents. Two observables are regarded as *equal* if and only if their typed variables are equal, they model the same feature and, in that context, one takes a given value if and only if the other does.

Being an abstraction, an observable is not necessarily meant to result from quantitative measurement or even empirical perception. The "feature of the system under consideration" might be a physical magnitude, but we shall see that it might also be an artefact of a conceptual model, constructed entirely for the purpose of analysis.

An observable, being a typed variable, has specifically determined possible values. In particular:

> Definition: An observable is called *discrete* if and only if its type has only finitely many possible values; otherwise it is called *analogue*.[4]

In this paper, we are interested in observables as a means of describing behaviour at a precisely qualified (though seldom numerical) level of abstraction; in general, several observables will be employed.

[4] The distinction is really a matter of topology rather than cardinality. However, this definition serves our present purposes.

2.3. Five Examples

A good way to gain some acquaintance with the previous concepts is by looking at a few simple examples.

1) Suppose Peter and Ann wish to study some physical human attributes. To do so Peter, in Oxford, introduces a variable, h, whose type consists of rational numbers. The typed variable h becomes an (analogue) observable once it is decided that the variable h represents the height of a person, using the Imperial system (feet and parts thereof). To explain the definition of equality of observables, suppose that Ann, in Rome, is also interested in observing human physical attributes, and defines the same typed variable but declares that it represents height in metres and parts thereof. Their *typed variables* are the same, but they differ as *observables*: for a given person, the two variables take different representing values. This example shows the importance of making clear the interpretation by which a typed variable becomes an observable.

2) Consider next an example of an *ill-typed variable*. Suppose we are interested in the roles played by people in some community; we could not introduce an observable standing for those beauticians who depilate just those people who do not depilate themselves, for it is well-known that such a variable would not be well typed (Russell [1902]). Similarly, each of the standard antinomies reflects an ill-typed variable (Hughes and Brecht [1976]). Of course, the modeller is at liberty to choose whatever type befits the application and, if that involves a potential antinomy, then the appropriate type might turn out to be a non-well-founded set (Barwise and Etchemendy [1987]). However, in this paper we shall operate entirely within the boundaries of standard naive set theory.

3) Gassendi provides another nice example, to which I shall return in the conclusion. As he wrote in his *Fifth Set of Objections to Descartes' Meditations* "If we are asking about wine, and looking for the kind of knowledge which is superior to common knowledge, it will hardly be enough for you to say 'wine is a liquid thing, which is compressed from grapes, white or red, sweet, intoxicating' and so on. You will have to attempt to investigate and somehow explain its internal substance, showing how it can be seen to

be manufactured from spirits, tartar, the distillate, and other ingredients mixed together in such and such quantities and proportions."

What Gassendi seems to have in mind is that observables relating to tasting wine include the attributes that commonly appear on "tasting sheets": *nose* (representing bouquet), *legs* or *tears* (viscosity), *robe* (peripheral colour), *colour*, *clarity*, *sweetness*, *acidity*, *fruit*, *tannicity*, *length* and so on, each with a determined type. If two wine tasters choose different types for, say, *colour* (as is usually the case) then the observables are different, despite the fact that their variables have the same name and represent the same feature in reality. Indeed, as they have different types they are not even equal as typed variables.

Information about how wine quality is perceived to vary with time – how the wine "ages" (Robinson [1989]) – is important for the running of a cellar. An appropriate observable is the typed variable a, which is a function associating to each year $y{:}\mathit{Years}$ a perceived quality $a(y){:}\mathit{Quality}$, where the types *Years* and *Quality* may be assumed to have been previously defined. Thus, a is a function from *Years* to *Quality*, written $a{:}\ \mathit{Time} \to \mathit{Quality}$. This example shows that, in general, types are constructed from more basic types, and that observables may correspond to operations, taking input and yielding output. Indeed, an observable may be of an arbitrarily complex type.

4) The definition of an observable reflects a particular view or attitude towards the entity being studied. Most commonly, it corresponds to a simplification, in which case nondeterminism, not exhibited by the entity itself, may arise. The method is successful when the entity can be understood by combining the simplifications. Let us consider another example.

In observing a game of chess, one would expect to record the moves of the game.[5] Other observables might include the time taken per move, the body language of the players, and so on. Suppose we are able to view a chessboard by just looking along *files* (the columns stretching from player to player). When we play "files-chess", we are

[5] As the reader probably knows, this is done by recording the history of the game: move by move the state of each piece on the board is recorded – in English algebraic notation – by rank and file, the piece being moved and the consequences of the move.

unable to see the ranks (the parallel rows between the players) or the individual squares. Files cannot sensibly be attributed a colour black or white, but each may be observed to be occupied by a set of pieces (namely those that appear along that file), identified in the usual way (king, queen and so forth). In "files-chess", a move may be observed by the effect it has on the file of the piece being moved. For example, a knight moves one or two files either left or right from its starting file; a bishop is indistinguishable from a rook, which moves along a rank; and a rook that moves along a file appears to remain stationary. Whether or not a move results in a piece being captured, appears to be nondeterministic. "Files-chess" seems to be an almost random game.

Whilst the "underlying" game is virtually impossible to reconstruct, each state of the game and each move (i.e., each operation on the state of the game) can be "tracked" within this dimensionally-impoverished family of observables. If one then takes a second view, corresponding instead to rank, we obtain "ranks-chess". Once the two views are combined, the original, bi-dimensional game of chess can be recovered, since each state is determined by its rank and file projections, for each move. The two disjoint observations together, namely "files-chess" + "ranks-chess", reveal the underlying game.

5) The degree to which a type is appropriate depends on its context and use. For example, to describe the state of a traffic light in Rome one might decide to consider an observable *colour* of type {*red*, *amber*, *green*} that corresponds to the colour indicated by the light. This option abstracts the length of time for which the particular colour has been displayed, the brightness of the light, the height of the traffic light, and so on. This is why the choice of type corresponds to a decision about how the phenomenon is to be regarded. To specify such a traffic light for the purpose of construction, a more appropriate type would comprise a numerical measure of wavelength (see section 2.6). Furthermore, if we are in Oxford, the type of colour would be a little more complex, since – in addition to red, amber and green – red and amber are displayed simultaneously for part of the cycle. So, an appropriate type would be {*red*, *amber*, *green*, *red-amber*}.

2.4. Level of Abstraction

We are now ready to appreciate the basic concept of *level of abstraction* (LoA).

Any collection of typed variables can, in principle, be combined into a single "vector" observable, whose type is the Cartesian product of the types of the constituent variables. In the wine example, the type *Quality* might be chosen to consist of the Cartesian product of the types *Nose*, *Robe*, *Colour*, *Acidity*, *Fruit* and *Length*. The result would be a single, more complex, observable. In practice, however, such vectorisation is unwieldy, since the expression of a constraint on just some of the observables would require a projection notation to single out those observables from the vector. Instead, I shall base our approach on a *collection* of observables, that is, a level of abstraction:

> Definition: A *level of abstraction* (*LoA*) is a finite but non-empty set of observables. No order is assigned to the observables, which are expected to be the building blocks in a theory characterised by their very definition. A LoA is called *discrete* (respectively *analogue*) if and only if all its observables are discrete (respectively analogue); otherwise it is called *hybrid*.

Consider the wine example. Different LoAs may be appropriate for different purposes. To evaluate a wine, the "tasting LoA", consisting of observables like those mentioned in the previous section, would be relevant. For the purpose of ordering wine, a "purchasing LoA" (containing observables like *maker*, *region*, *vintage*, *supplier*, *quantity*, *price*, and so on) would be appropriate; but here the "tasting LoA" would be irrelevant. For the purpose of storing and serving wine – the "cellaring LoA" (containing observables for *maker*, *type of wine*, *drinking window*, *serving temperature*, *decanting time*, *alcohol level*, *food matchings*, *quantity remaining in the cellar*, and so on) would be relevant.

The traditional sciences tend to be dominated by analogue LoAs, the humanities and information science by discrete LoAs and mathematics by hybrid LoAs. We are about to see why the resulting theories are fundamentally different.

2.5. Behaviour

The definition of observables is only the first step in studying a system at a given LoA. The second step consists in deciding what relationships hold between the observables. This, in turn, requires the introduction of the concept of system "behaviour". We shall see that it is the fundamentally different ways of describing behaviour in analogue and discrete systems that account for the differences in the resulting theories.

Not all values exhibited by combinations of observables in a LoA may be realised by the system being modelled. For example, if the four traffic lights at an intersection are modelled by four observables, each representing the colour of a light, the lights cannot in fact all be green together (assuming they work properly). In other words, the combination in which each observable is green cannot be realised in the system being modelled, although the types chosen allow it. Similarly, the choice of types corresponding to a rank-and-file description of a game of chess allows any piece to be placed on any square, but in the actual game two pieces cannot occupy the same square simultaneously.

Some technique is therefore required to describe those combinations of observable values that are actually acceptable. The most general method is simply to describe all the allowed combinations of values. Such a description is determined by a predicate, whose allowed combinations of values is called the "system behaviours".

> Definition: the *behaviour* of a system, at a given LoA, is defined to consist of a predicate whose free variables are observables at that LoA. The substitutions of values for observables that make the predicate true are called the *system behaviours*. A *moderated LoA* is defined to consist of a LoA together with a behaviour at that LoA.

Consider two previous examples. In reality, human height does not take arbitrary rational values, for it is always positive and bounded above by (say) nine feet. The variable h, representing height, is therefore constrained to reflect reality by defining its behaviour to consist of the predicate $0 < h < 9$, in which case any value of h in that interval is a "system" behaviour. Likewise, wine too is not realistically described by arbitrary combinations of the aforementioned observables. For instance, it cannot be both white and highly tannic.

Since Newton and Leibniz, the behaviours of analogue observables, studied in science, have typically been described by differential equations. A small change in one observable results in a small, quantified change in the overall system behaviour. Accordingly, it is the rates at which those smooth observables vary which is most conveniently described.[6] The desired behaviour of the system then consists of the solution of the differential equations. However, this is a special case of a predicate: the predicate holds at just those values satisfying the differential equation. If a complex system is approximated by simpler systems, then the differential calculus provides a supporting method for quantifying the approximation.

The use of predicates to demarcate system behaviour is essential in any (nontrivial) analysis of discrete systems because in the latter no such continuity holds: the change of an observable by a single value may result in a radical and arbitrary change in system behaviour. Yet, complexity demands some kind of comprehension of the system in terms of simple approximations. When this is possible, the approximating behaviours are described exactly, by a predicate, at a given LoA, and it is the LoAs that vary, becoming more comprehensive and embracing more detailed behaviours, until the final LoA accounts for the desired behaviours. Thus, the formalism provided by the method of abstraction can be seen as doing for discrete systems what differential calculus has traditionally done for analogue systems.

Likewise, the use of predicates is essential in subjects like information and computer science, where discrete observables are paramount and hence predicates are required to describe a system behaviour. In particular, state-based methods like *Z* (Hayes and Flinn [1993], Spivey [1992]) provide a notation for structuring complex observables and behaviours in terms of simpler ones. Their primary concern is with the syntax for

[6] It is interesting to note that the catastrophes of *chaos theory* are not smooth; although they do appear so when extra observables are added, taking the behaviour into a smooth curve on a higher-dimensional manifold. Typically, chaotic models are weaker than traditional models, their observables merely reflecting *average* or *long-term* behaviour. The nature of the models is clarified by making explicit the

expressing those predicates, an issue that will be avoided in this paper by stating predicates informally.

The time has now come to combine approximating, moderated LoAs to form the primary concept of the method of abstraction.

2.6. Gradient of Abstraction

For a given (empirical or conceptual) system or feature, different LoAs correspond to different representations or views. A *Gradient of Abstractions* (GoA) is a formalism defined to facilitate discussion of discrete systems over a range of LoAs. Whilst a LoA formalises the scope or granularity of a single model, a GoA provides a way of varying the LoA in order to make observations at differing levels of abstraction.

For example, in evaluating wine one might be interested in the GoA consisting of the "tasting" and "purchasing" LoAs, whilst in managing a cellar one might be interested in the GoA consisting of the "cellaring" LoA together with a sequence of annual results of observation using the "tasting" LoA. The reader acquainted with Dennett's idea of "stances" may compare them to a GoA (more on this in section four).

In general, the observations at each LoA must be explicitly related to those at the others; to do so, one uses a family of relations between the LoAs. For this, I need to recall some (standard) preliminary notation.

Notation: A *relation* R from a set A to a set C is a subset of the Cartesian product $A \times C$. R is thought of as relating just those pairs (a, c) that belong to the relation. The *reverse* of R is its mirror image: $\{(c, a) \mid (a, c) \in R\}$. A relation R from A to C translates any predicate p on A to the predicate $P_R(p)$ on C that holds at just those $c{:}C$, which are the image through R of some $a{:}A$ satisfying p

$$P_R(p)(c) = \exists a{:}\ A\ \ R(a,c) \wedge p(a)$$

LoA.

We have finally come to the main definition of the paper:

> Definition: A *gradient of abstractions*, *GoA*, is defined to consist of a finite set[7] $\{L_i \mid 0 \leq i < n\}$ of moderated LoAs L_i, a family of relations $R_{i,j} \subseteq L_i \times L_j$, for $0 \leq i \neq j < n$, relating the observables of each pair L_i and L_j of distinct LoAs in such a way that:
>
> 1. the relationships are inverse: for $i \neq j$, $R_{i,j}$ is the reverse of $R_{j,i}$
> 2. the behaviour p_j at L_j is at least as strong as the translated behaviour
>
> $$P_{R_{i,j}}(p_i)\, p_j \Rightarrow P_{R_{i,j}}(p_i). \qquad (1)$$
>
> and for each interpreted type $x{:}X$ and $y{:}Y$ in L_i and $\mathrm{L}j$ respectively, such that $(x{:}X, y{:}Y)$ is in R_{ij}, a relation $\mathrm{R}xy \subset \mathrm{X} \times \mathrm{Y}$.[8]

Two GoAs are regarded as *equal* if and only if they have the same moderated LoAs (i.e., the same LoAs and moderating behaviours) and their families of relations are equal. A GoA is called *discrete* if and only if all its constituent LoAs are discrete.

Condition (1) means that the behaviour moderating each lower LoA is *consistent* with that specified by a higher LoA. Without it, the behaviours of the various LoAs constituting a GoA would have no connection with each other. A special case, to be elaborated below in the definition of "nestedness", helps to clarify the point.

If one LoA L_i extends another L_j by adding new observables, then the relation $R_{i,j}$ is the inclusion of the observables of L_i in those of L_j and (1) reduces to this: the constraints imposed on the observables at LoA L_i remain true at LoA L_j, where "new" observables lie outside the range of $R_{i,j}$.

A GoA whose sequence contains just one element evidently reduces to a single LoA. So our definition of "LoA" is subsumed by that of "GoA".

[7] The case of infinite sets has application to analogue systems but is not considered here.

[8] I wish to thank Jesse F. Hughes for having pointed out to me the last requirement, without which only the variables would be related but not the elements of their types.

The consistency conditions imposed by the relations $R_{i,j}$ are in general quite weak. It is possible, though of little help in practice, to define GoAs in which the relations connect the LoAs cyclically. Of much more use are the following two important kinds of GoA: "disjoint" GoAs (whose views are complementary) and "nested" GoAs (whose views provide successively more information). Before defining them, some further notations need to be introduced.

It will be recalled that a *function f* from a set C to a set A is a relation, i.e., a subset of the Cartesian product $C \times A$, which is single-valued, that is:

$$\forall c{:}C \;\; \forall a, a'{:}A \;\; ((c,a) \in f \wedge (c,a') \in f) \Rightarrow a = a'$$

this means that the notation $f(c) = a$ is a well-defined alternative to $(c,a) \in f)$, and total, that is:

$$\forall c{:}C \;\; \exists a{:}A \;\; f(c) = a$$

this means that $f(c)$ is defined for each $c{:}C$. A function is then called *surjective* if and only if every element in the target set lies in the range of the function, that is:

$$\forall a{:}A \;\; \exists c{:}C \;\; f(c) = a.$$

We are now ready to introduce the definition of GoA:

> Definition: A GoA is called *disjoint* if and only if the L_i are pairwise disjoint (i.e., taken two at a time, they have no observable in common) and the relations are all empty. It is called *nested* if and only if the only nonempty relations are those between L_i and L_{i+1}, for each $0 \leq i < n-1$, and moreover the reverse of each $R_{i,\, i+1}$ is a surjective function from the observables of L_{i+1} to those of L_i.

A disjoint GoA is chosen to describe a system as the combination of several non-overlapping components. This is useful when different aspects of the system behaviour are better modelled as being determined by the values of distinct observables. Think for example of a typical case of Cartesian dualism, in which a disjoint GoA models the brain and its observables as a *res extensa* and the mind and its observables as a *res cogitans*. The case of a disjoint GoA is rather simple, since the LoAs are more typically tied together by common observations. For example, the services in a domestic dwelling may be represented by LoAs for electricity, plumbing, telephone, security and gas. Without going into detail about the constituent observables, it is easy to see that, in an accurate representation, the electrical and plumbing LoAs would overlap whilst the telephone and plumbing would not. Following the philosophical example, this would correspond to a case in which some form of epiphenomenalism is being supported.

A nested GoA (see Figure1) is chosen to describe a complex system exactly at each level of abstraction and incrementally more accurately. The condition that the functions be surjective means that any abstract observation has at least one concrete counterpart. As a result, the translation functions cannot overlook any behaviour at an abstract LoA: behaviours lying outside the range of a function translate to the predicate *false*. The condition that the reversed relations be functions means that each observation at a concrete LoA comes from at most one observation at a more abstract LoA (although the converse fails in general, allowing one abstract observable to be refined by many concrete observables). As a result the translation functions become simpler.

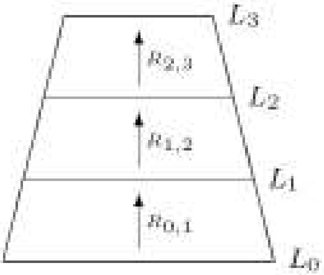

Figure 1 Nested GoA with four Levels of Abstraction

Using the previous example regarding the brain, ideally neuroscientific studies rely on nested GoAs, as they proceed from the investigation of whole brain functions and related areas, such as specific kinds of memories, to investigations of the physiological basis of memory storage in neurons. On a more prosaic note, let me recall the case of a traffic light, which is observed to have colour *colour* of type {*red*, *amber*, *green*}. This is captured by a LoA, L_0, having that single observable. If one wishes to be more precise about colour, e.g. for the purpose of constructing a new traffic light, one might consider a second LoA, L_1, having the variable *wl* whose type is a positive real number corresponding to the wavelength of the colour. To determine the behaviour of L_1, Suppose that constants $\lambda_{red} < \lambda_{red'}$ delimit the wavelength of red, and similarly for amber and green. Then the behaviour of L_1 is simply this predicate with free variable *wl*:

$$(\lambda_{red} \leq wl \leq \lambda_{red'}) \vee (\lambda_{amber} \leq wl \leq \lambda_{amber'}) \vee (\lambda_{green} \leq wl \leq \lambda_{green'}).$$

The sequence consisting of the LoA L_0 and the moderated LoA L_1 forms a nested GoA. Intuitively, the smaller, abstract, type {*red*, *amber*, *green*} is a projection of the larger. The relevant relation associates to each value *c*:{*red*, *amber*, *green*} a band of wavelengths perceived as that colour. Formally, *R*(*colour*,*wl*) is defined to hold if and only if, for each *c*:{*red*, *amber*, *green*}:

$$colour = c \quad \leftrightarrow \quad \lambda_c \leq wl \leq \lambda_{c'}.$$

In the wine example, the first LoA might be defined to consist of the variable “kind” having type consisting of *red*, *white*, *rose* under the obvious representation. A second LoA might be defined to consist of the observable “kind” having type:

$$\{stillred, sparklingred, stillwhite, sparklingwhite, stillrose, sparklingrose\}.$$

Although the second type does not contain the first, it produces greater resolution under the obvious projection relation. Thus, the GoA consisting of those two LoAs is nested.

These two important forms of GoA – disjoint and nested – are in fact interchangeable, at least theoretically. For if A and B are disjoint sets then A and their union $A \cup B$ are increasing sets and the former is embedded in the latter. Thus, a disjoint GoA can be converted to a nested one. Conversely, if A and B are increasing sets with the former embedded in the latter, then A and the set difference $A \setminus B$ are disjoint sets. Thus, a nested GoA can be converted to a disjoint one.

Following the technique used to define a nested GoA, it is possible to define less restricted but still hierarchical GoAs. Important examples include tree-like structures, of which our nested GoAs are a special, linear case.

For theoretical purposes, the information captured in a GoA can be expressed equivalently as a single LoA of a more complicated type, namely one whose single LoA has a type equal to the sequence of the LoAs of the complex interface. However, the current definition is better suited to application.

2.7. The Method of Abstraction

Models are the outcome of an analysis of a system, developed at some LoA(s) for some purpose. An important contribution of these ideas is to make precise the commitment to a LoA/GoA before further elaborating a theory. This is called the *method of abstraction*. Four advantages of the method can be highlighted here.

First, and most importantly for our present concerns, it is useful to specify the meaning of "indirect knowledge"[9] in terms of knowledge mediated by a LoA.

It follows, (second advantage) that specifying the LoA means clarifying, from the outset, the range of questions that (a) can be meaningfully asked and (b) are answerable in principle. One might think of the input of a LoA as consisting of the system under analysis, comprising a set of *data*; its output is a *model* of the system, comprising *information*. The quantity of information in a model varies with the LoA: a lower LoA, of greater resolution or finer granularity, produces a model that contains

[9] Direct knowledge is to be understood here as typically knowledge of one's mental states, which is apparently not mediated; indirect knowledge is usually taken to be knowledge that is obtained inferentially or through some other form of mediated communication with the world.

more information than a model produced at a higher, or more abstract, LoA. Thus, a given LoA provides a quantified commitment to the kind and amount of information that can be "extracted" from the system. The choice of a LoA pre-determines the type and quantity of data that can be considered and hence the information that can be contained in the model. So, knowing at which LoA the system is being analysed is indispensable, for it means knowing the scope and limits of the model being developed.

Third, being explicit about the LoA adopted provides a healthy antidote to ambiguities, equivocations and other fallacies or errors due to level-shifting, such as Aristotle's "metabasis eis allo genos" (shifting from one genus to another), Ryle's "category-mistakes", and Kant's "antinomies of pure reason".

Fourth, by stating its LoA, a theory is forced to make explicit and clarify its ontological commitment, in the following way.

We have seen that a model is the output of the analysis of a system, developed at some LoA(s), for some purpose. So a theory of a system comprises at least three components:

i) a LoA, which determines the range of available observables and allows the theory to investigate the system under analysis and to elaborate

ii) the ensuing model of that system, which identifies

iii) a structure of the system at the given LoA.

Let us refer to this as the system-level-model-structure (or SLMS) scheme (see Fig. 1).

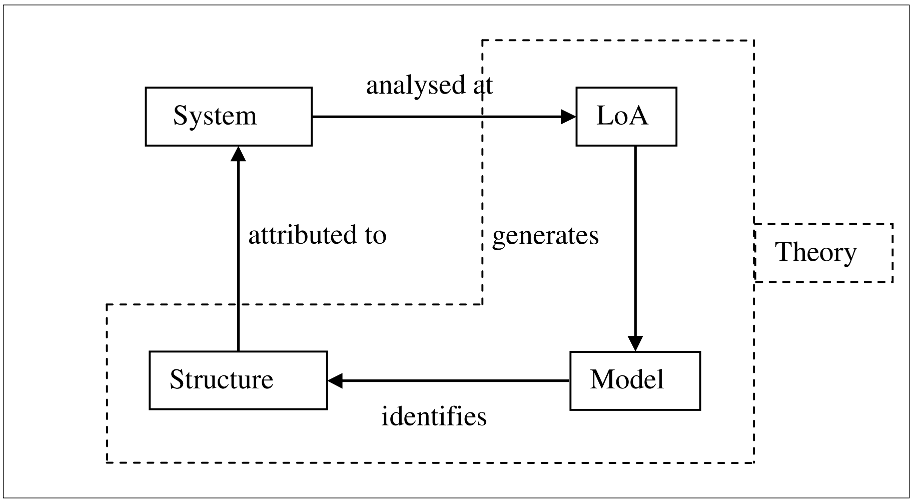

Fig. 1: the SLMS scheme

The ontological commitment of a theory can be clearly understood by distinguishing between a *committing* and a *committed* component, within the SLMS scheme.

A theory commits itself ontologically by opting for a specific LoA, whose application commits the theory to a particular model of the system. The order is purely logical. By adopting a LoA, the theory decides what kind of observables are going to play a role in elaborating the model. In our traffic light example, suppose the LoA commits the theory to take into account only data relative to colour type. When the LoA generates a model, i.e. when the observables are instantiated, the theory is committed to a particular view of the system. Again, in our example, this might be the specific colours used in the model.

To summarise, by accepting a LoA a theory commits itself to the existence of certain types of objects, the types constituting the LoA (by trying to model a traffic light in terms of three colours one shows one's commitment to the existence of a traffic light of that kind, i.e. one that could be found in Rome, but not in Oxford), while by endorsing the ensuing models the theory commits itself to the corresponding tokens (by endorsing a particular model, which is the outcome of the interpretation of the data at the chosen LoA, one commits oneself to that model, e.g. one now cannot have a fourth phase when amber and green are on at the same time). Figure 2 summarises this distinction (note that, for the sake of simplicity the term "theory" is the dotted line that comprises, as above, LoA, model and structure).

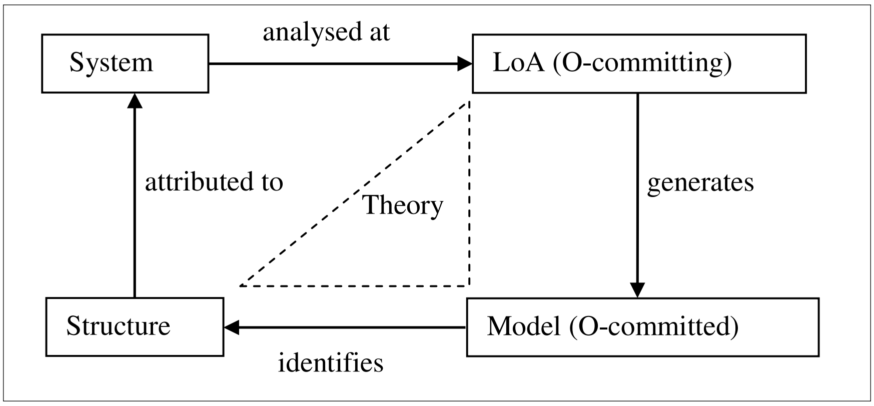

Fig. 2: the SLMS scheme with ontological commitment

3. A Classic Application of the Method of Abstraction

A simple way to introduce the method of levels of abstraction (LoAs) and highlight its philosophical importance is by showing how closely it resembles Kant's transcendental approach. The resemblance is not casual, but a scholarly explanation of this "family relation" would take us too far, in an exegetical direction that I am not interested in pursuing here. Rather, it is interesting to highlight here the similarities between the two methods by referring to Kant's classic discussion of the "antinomies of pure reason". The only point that the reader may wish to keep in mind, lest I give the impression that Kant gets away too lightly with his transcendentalism, is that, in Kant, knowledge of reality is indirect because of the mind's transcendental schematism but, after the downfall of Neo-Kantism and Cassirer's and C. I. Lewis' revisions of the transcendental, an approach is needed that is less infra-subjective, mental (if not psychologistic), innatist, individualistic and rigid. In Floridi and Sanders [2004c] and Floridi and Sanders [2004b], the *method of levels of abstraction* has been proposed as a more inter-subjective, socially constructible (hence possibly conventional), dynamic and flexible way to further Kant's approach. This is a step away from *internal realism* (kinds, categories and structures of the world are only a function of our conceptual schemes), but not yet a step into *external* or *metaphysical realism* (kinds, categories and structures of the world belong to the world and are not a function of our conceptual schemes, either causally or ontologically). If necessary, it might be called *liminal realism*, for reasons that will become clearer below. With this clarification in the background, let us now see the similarities.

As is well-known, each of the four antinomies comprises a thesis and an antithesis, which are supposed to be both reasonable and irreconcilable. I list them here by slightly adapting their formulation from Kant's *Critique of Pure Reason* (Kant [1998]):

1) Thesis: the world is finite; it has a beginning in time and is limited in space.

   Antithesis: the world is infinite, it has no beginning in time and no limit in space (A 426-7/B 454-5).

2) Thesis: the world is discrete; everything in the world consists of elements that are ultimately simple and hence indivisible.

   Antithesis: the world is continuous; nothing in the world is simple, but everything is composite and hence infinitely divisible (A 434-5/B 462-3).

3) Thesis: there is freedom; to explain causal events in the world it is necessary to refer both to the laws of nature and to freedom.

   Antithesis: there is no freedom; everything that happens in the world occurs only in accordance with natural causation (A 444-5/B 462-3).

4) Thesis: there is in the world an absolutely necessary being.

   Antithesis: there is nothing necessary in the world, but everything is contingent (A 452-3/B 480-1).

What I wish to stress here is that Kant's transcendental method and the method of abstraction converge both on the evaluation and on the resolution of these antinomies.

As Kant argues, the conflict is not between empirical experience and logical analysis. Rather, the four antinomies are generated by an unconstrained demand for unconditioned answers to fundamental problems concerning (1) time and space, (2) complexity/granularity, (3) causality and freedom or (4) modality. And here is where my assessment agrees with Kant's: the strive for something unconditioned is equivalent to the natural yet profoundly mistaken attempt to analyse a system (the world in itself, for Kant, but it could also be a more limited system) independently of any (specification of) the level of abstraction at which the analysis is being conducted, the questions are being posed and the answers are being offered. In other words, trying to overstep the limits set by the LoA leads to a conceptual mess.

As for the resolution, Kant divides the antinomies into two groups. He then shows that, in the first two antinomies, both the thesis and the antithesis are untenable because the search for the unconditioned mistakes time and space, and complexity/granularity, for features of the system instead of realising that they are properties set by (or constituting) the level of abstraction at which the system is investigated and hence, as such, subject to alternative formatting. Following Kant, one may say that, assuming for the sake of simplicity that a LoA is comparable to an interface, it makes no sense to wonder whether the system under observation is finite in time, space and granularity in itself, independently of the LoA at which it is being analysed, since this is a feature of the interface, and different interfaces may be adopted depending on needs and requirements. So, from a LoA approach, I agree with Kant: neither the thesis nor the antithesis in (1) and (2) are tenable.

Regarding the third and fourth antinomy, Kant argues that both the thesis and the antithesis might be tenable, thus coming close to what has been defined above as a disjoint GoA. The mistake here lies in confusing what qualifies the phenomenal world of experience – which relies on causal relations and is characterised by contingency – with what might qualify the noumenal world of things in themselves – which may include freedom and necessary existence, but that remains inaccessible through experience. In the language of the method of abstraction, this means that models, i.e., the outcomes of the analyses of systems, are always characterised by natural laws of causality and a modality of contingencies, but this does not disprove the existence of freedom and God "in the systems", two issues with respect to which one may remain agnostic and uncommitted.

All this clarifies three important aspects of the method of abstraction. First, the method is Kantian in nature. Although it does not inherit from Kant any mental or subject-based feature, it is a transcendental approach, which considers the conditions of possibility of the analysis (experience) of a particular system.

Second, the method is anti-metaphysical, again in a Kantian sense. Metaphysics is – when used as a negative label – what is done by sloppy reasoning when it pretends to develop a theory without taking into consideration, at least implicitly, the level of abstraction at which it is being developed. In other words, metaphysics is that LoA-free zone where anyone can say anything without fear of ever being proved wrong, as long as the basic law of non-contradiction is respected. Such an unconstrained game of ideas should be found dull and frustrating by anyone genuinely interested in knowledge.

Third, the method provides a powerful tool to approach significant issues in philosophy. We have just seen how it can dispose of false antinomies in a Kantian way. I shall mention a few more examples in the conclusion.

4. The Philosophy of the Method of Abstraction

The time has come to provide further conceptual clarification concerning the nature and consequences of the method of abstraction. In this section, I relate the relevant work of Marr, Pylyshyn, Dennett and Davidson to the method of abstraction, and discuss the thorny issues of relativism and antirealism. A word of warning may be in order. When confronted with a new theory or method, it is natural to compare it and perhaps (mistakenly) identify it with something old and well-established. In particular, previous theories or methods can work as powerful magnets that end by attracting anything that comes close to their space of influence, blurring all differences. So this section aims at putting some distance between some old acquaintances and the new proposal.

4.1. Levels of Organization and of Explanation

Several important ways have been proposed for speaking of the levels of analysis of a system. The following two families can be singled out as most representative:

1) *Levels of organization* (LoOs) support an *ontological* approach, according to which the system under analysis is supposed to have a (usually hierarchical) structure in itself, or *de re*, which is allegedly uncovered by its description and objectively formulated in some neutral observation language (Newell [1990], Simon [1996]). For example, levels of communication, of decision processing (Mesarovic et al. [1970]) and of information flow can all be presented as specific instances of analysis in terms of LoOs.

There is a twofold connection between LoOs and LoAs. If the hierarchical structure of the system itself is thought of as a GoA, then for each constituent LoA there is a corresponding LoO. Alternatively, one can conceive the analysis of the system, not the system itself, as being the object of study. Then the method of abstraction leads one to consider a GoA whose constituent LoAs are the LoOs. Note that, since the system under analysis may be an artefact, knowledge of its LoO may be available constructively, i.e., in terms of knowledge of its specifications.

2) *Levels of explanation* (LoEs) support an *epistemological* approach, quite common in cognitive and computer science (Benjamin et al. [1998]). Strictly speaking, the LoEs do not really pertain to the system or its model. They provide a way to distinguish between different epistemic approaches and goals, such as when one analyses an exam question from the students' or the teacher's perspectives, or the description of the functions of a technological artefact from the designer's, the user's, the expert's or the layperson's point of view.

A LoE is an important kind of LoA. It is pragmatic and makes no pretence of reflecting an ultimate description of the system. It has been defined with a specific practical view or use in mind. Manuals, pitched at the inexpert user, indicating "how to" with no idea of "why", provide a good example.

The two kinds of "structured analysis" just introduced are of course interrelated. Different LoEs – e.g., the end-user's LoE of how an applications package is to be used versus the programmer's LoE of how it is executed by the machine – are connected with different LoAs – e.g., the end-user's LoA represented by a specific graphic interface versus the programmer's code – which in turn are connected with different LoO – e.g., the commonsensical WYSIWYG versus the software architecture. However, LoAs provide a foundation for both, and LoOs, LoEs and LoAs should not be confused. Let us consider some clarifying examples.

One of the most interesting and influential cases of multi-layered analysis is provided by Marr's three-levels hypothesis. After Marr [1982], it has become common in cognitive and philosophical studies (McClamrock [1991]) to assume that a reasonably complex system can be understood only by distinguishing between levels of analysis.

Here is how Marr himself put it: "Almost never can a complex system of any kind be understood as a simple extrapolation from the properties of its elementary components. Consider for example, some gas in a bottle. A description of thermodynamic effects – temperature, pressure, density, and the relationships among these factors – is not formulated by using a large set of equations, one for each of the particles involved. Such effects are described at their own level, that of an enormous collection of particles; the effort is to show that in principle the microscopic and the macroscopic descriptions are consistent with one another. If one hopes to achieve a full understanding of a system as complicated as a nervous system, a developing embryo, a set of metabolic pathways, a bottle of gas, or even a large computer program, then one must be prepared to contemplate different kinds of explanation at different levels of description that are linked, at least in principle, into a cohesive whole, even if linking the levels in complete detail is impractical. For the specific case of a system that solves an information-processing problem, there are in addition the twin strands of process and representation, and both these ideas need some discussion." (Marr [1982], pp. 19–20).

In particular, in the case of an information-processing system, Marr and his followers suggest the adoption of three levels of analysis (all the following quotations are from Marr [1982]):

1) *the computational level*. This is a description of "the abstract computational theory of the device, in which the performance of the device is characterised as a mapping from one kind of information structures, the abstract properties of this mapping are defined precisely, and its appropriateness and adequacy for the task at hand are demonstrated" (p. 24);
2) *the algorithmic level*. This is a description of "the choice of representation for the input and output and the algorithm to be used to transform one into the other" (p. 24-25);
3) *the implementational level*. This is a description of "the details of how the algorithm and representation are realized physically – the detailed computer architecture, so to speak." (p. 25).
The three levels are supposed to be loosely connected and in a one-to-many mapping relation: for any computational description of a particular information-processing problem there may be several algorithms for solving that problem, and any algorithm may be implemented in several ways.

Along similar lines, Pylyshyn [1984] has spoken of the *semantic*, the *syntactic*, and the *physical levels of description* of an information-processing system, with the (level of) *functional architecture* of the system playing the role of a bridge between Marr's algorithmic and implementational levels. And Dennett [1987] has proposed a hierarchical model of explanation based on three different "stances": the intentional stance, according to which the system is treated, for explanatory purposes, *as if* it were a rational, thinking agent attempting to carry out a particular task successfully; the design stance, which concerns the general principles governing the design of any system that might carry out those tasks successfully; and the physical stance, which considers how a system implementing the appropriate design-level principles might be physically constructed.

The tripartite approaches of Marr, Pylyshyn and Dennett share three important features. First, they are each readily formalised in terms of GoAs with three LoAs. Second, they do not distinguish between LoO, LoE and LoA; and this because (third feature) they assign a privileged role to explanations. As a result, their ontological commitment is embedded and hence concealed. The common reasoning seems to be the following: "this is the right level of analysis because that is the right LoO", where no justification is offered for why that LoO is chosen as the right one. Nor is the epistemological commitment made explicit or defended; it is merely presupposed. This is where the method of abstraction provides a significant advantage. By starting from a clear endorsement of each specific LoA, a strong and conscious effort can be made to uncover the ontological commitment of a theory (and hence of a set of explanations), which now needs explicit acceptance on the part of the user, and requires no hidden epistemological commitment, which now can explicitly vary depending on goals and requirements.

4.2. Conceptual Schemes

The resemblance between LoAs and *conceptual schemes* (CSs) is close enough to require further clarification. In this section, I shall briefly compare the two. The aim is not to provide an exegetical interpretation or a philosophical analysis of Davidson's famous criticism of the possibility of irreducible CSs, but rather to clarify further the nature of LoAs and explain why LoAs can be irreducible, although in a sense different from that preferred by supporters of the irreducibility of CSs.[10]

According to Davidson, all CSs share four features (the following quotations are from Davidson [1974]):

1) CSs are clusters or networks of (possibly acquired) categories. "Conceptual schemes, we are told, are ways of organizing experience; they are systems of categories that give form to the data of sensation; they are points of view from which individuals, cultures, or periods survey the passing scene" (p. 183).

2) CSs describe or organise the world or its experience for communities of speakers. "Conceptual schemes (languages) either organize something, or they fit it", and as "for

[10] Newell reached similar conclusions, despite the fact that he treated LoA as LoO, an ontological form of levelism that allowed him to escape relativism and antirealism more easily, see Newell [1982] and Newell [1993].

the entities that get organized, or which the scheme must fit, I think again we may detect two main ideas: either it is reality (the universe, the world, nature), or it is experience (the passing show, surface irritations, sensory promptings, sense-data, the given)" (p. 192).

3) CSs are inescapable, in the sense that communities of speakers are entrapped within their CSs.

4) CSs are not intertranslatable.

Davidson argues against the existence of CSs as inescapable (from within) and impenetrable (from without) ways of looking at the world by interpreting CSs linguistically and then by trying to show that feature (4) is untenable. Could the strategy be exported to contrast the existence of equally inescapable and impenetrable LoAs? Not quite.

Let us examine what happens to the four features above when LoAs are in question:

a) LoAs are clusters or networks of observables. Since they deal with observables, LoAs are not an anthropocentric prerogative but allow a more general (or indeed less biased) approach. We do not have to limit ourselves to human beings or to communities of speakers. Different sorts of empirical or abstract agents – not only human beings but also computers, animals, plants, scientific theories, measurement instruments etc. – operate and deal with the world (or, better, with the data they glean from it) at some LoAs. By neatly decoupling LoAs from the agents that implement or use them, we avoid confusion between CSs, the languages in which they are formulated or embodied, and the agents that use them. I shall return to this point presently.

b) LoAs model the world or its experience. LoAs are anchored to their data, in the sense that they are constrained by them; they do not merely describe or organise them, they actually build models out of them. So the relation between models and their references (the analysed systems) is neither one of discovery, as in Davidson's CSs, nor one of invention, but one of design, to use an equally general category. It follows that, contrary to Davidson's CSs, it makes no sense to speak of LoAs as Xerox machines or personal organisers of some commonly shared ontology (the world or its experience). Ontological commitments are initially negotiated through the choice and shaping of LoAs, which therefore cannot presuppose a metaphysical omniscience.

Because of the differences between (1)–(2) and (a)–(b), the remaining two features acquire a significantly different meaning, when speaking of LoAs. Here is how the problem is reformulated. LoAs generate, and commit the agent to, information spaces. In holding that some LoAs can be irreducible and hence untranslatable I am not arguing that:

i) agents using LoAs can never move seamlessly from one information space to another. This is false. They obviously can, at least in some cases: just imagine gradually replacing some observables in the LoAs of an agent. This is equivalent to arguing that human beings cannot learn different languages. Note, however, that some agents may have their LoAs hardwired: imagine, for example, a thermometer;

ii) agents using LoAs can never expand their information spaces. This is also false. Given the nested nature of some LoAs and the possibility of constructing supersets of sets of observables, agents can aggregate increasingly large information spaces. This is equivalent to arguing that human speakers cannot expand their languages semantically, another obvious nonsense.

So, if we are talking about the agents using or implementing the LoAs, we know that agents can sometimes modify, expand or replace their LoAs, and hence some degree of intertranslatability, understood as the acquisition or evolution of new LoAs, is guaranteed. The point in question is another one, however, and concerns the relation between the LoAs themselves.

LoAs are the place at which (diverse) independent systems meet and act on or communicate with each other. If one reads carefully, one will notice that this is the definition of an interface. The systems interfaced may adapt or evolve their interfaces or adopt other interfaces, as in (i) and (ii), yet different interfaces may still remain mutually untranslatable. Consider, for example, the “tasting LoA” and the “purchasing LoA” in our wine example. But if two LoAs are untranslatable, it becomes perfectly reasonable to assume that:

iii) agents may inhabit only some types of information spaces in principle.

Some information spaces may remain inaccessible not just in practice but also in principle, or they may be accessible only asymmetrically, to some agents. Not only that, but given the variety of agents, what is accessible to one, or some, may not be accessible to all. This is easily explained in terms of modal logic and possible worlds understood as information spaces. The information space of a child is asymmetrically accessible from the information space of an adult, but the information space of a bat overlaps insufficiently with the information space of any human agent to guarantee a decent degree of translatability (Nagel [1974]).

In principle, some information spaces may remain forever disjoint from any other information spaces that some agents may be able to inhabit. When universalised, this is Kant's view of the noumenal world, which is accessible only to its creator. Does this imply that, after all, we are able to say what a radically inaccessible information space would be like, thus contradicting ourselves? Of course not. We are only pointing in the direction of the ineffable, without grasping it.

To return to Davidson, even conceding that he may be successful in criticising the concept of CSs, his arguments do not affect LoAs. The problem is that Davidson limits his consideration to information spaces that he assumes, without much reason, to be already linguistically and ontologically delimited. When this is the case, one may concede his point. However, LoAs do not vouch for the kind of epistemic realism, verificationism, panlinguism and representationist view of knowledge that Davidson implicitly assumes in analysing CSs. And once these fundamental assumptions are eliminated, Davidson's argument loses most of its strength. Incommensurable and untranslatable LoAs are perfectly possible, although we shall see that this provides no good ground for a defence of some form of radical conceptual relativism (section 4.3) or anti-realism (section 4.4).

Davidson’s criticism ends by shaping an optimistic approach to the problem of the incommensurability of scientific theories that supporters of the method of abstraction cannot share, but then, what conclusions can be drawn, from our analysis of LoAs, about the anti-realist reading of the history of science? An unqualified answer would fall victim to the same fallacy of un-layered abstraction I have been denouncing in the previous pages. The unexciting truth is that different episodes in the history of science are more or less comparable depending on the LoA adopted. Consider the great variety of building materials, requirements, conditions, needs and so on, which determine the actual features of a building. Does it make sense to compare a ranch house, a colonial home, a town house, a detached house, a semidetached house, a terraced house, a cottage, a thatched cottage, a country cottage, a flat in a single-storey building, and a Tuscan villa? The question cannot be sensibly answered unless one specifies the LoA at which the comparison is to be conducted. Likewise, my answer concerning the reading of the history of science is: given the nature of LoAs, it is always possible to formulate a LoA at which comparing different episodes in the history of science makes perfect sense. But do not ask absolute questions, for they just create an absolute mess.

4.3. Pluralism without Relativism

A LoA qualifies the level at which a system is considered. In this paper, I have argued that it must be made clear before the properties of the system can be sensibly discussed. In general, it seems that many disagreements might be clarified and resolved if the various “parties” make explicit their LoA. By structuring the explanandum, LoAs can reconcile the explanans. Yet, another crucial clarification is now in order. It must be stressed that a clear indication of the LoA at which a system is being analysed allows pluralism without falling into relativism or “perspectivism”, a term coined by Hales and Welshon [2000] in connection with Nietzsche’s philosophy. As remarked above, it would be a mistake to think that “anything goes” as long as one makes the LoA explicit, because LoAs can be mutually comparable and assessable, in terms of inter-LoA coherence, of their capacity to take full advantage of the same data and of their degree of fulfilment of the explanatory and predictive requirements laid down by the level of explanation. Thus, introducing an explicit reference to the LoA makes it clear that the model of a system is a function of the available observables, and that it is reasonable to rank different LoAs and to compare and assess the corresponding models.

4.4. Realism without Descriptivism

For a typed variable to be an observable it must be interpreted, a correspondence that has inevitably been left informal. This interpretation cannot be omitted: a LoA composed of typed variables called simply *x*, *y*, *z* and so on and treated rather formally, would leave the reader (or the writer some time later) with no hint of its domain of application. Whilst that is the benefit of mathematics, enabling its results to be applied whenever its axioms hold, in the method of abstraction it confers only obscurity. Does the informality of such an interpretation hint at some hidden circularity or infinite regress? Given the distinction between LoO and LoA, and the fact that there is no immediate access to any LoO that is LoA-free, how can an observable be defined as "realistic"? That is, must the system under consideration already be observed before a "realistic" observation can be defined? The mathematics underlying our definitions of typed variable and behaviour has been indicated (even if it is not always fully used in practice) to make the point that, in principle, the ingredients in a LoA can be formalised. There is no circularity: the heuristically appreciated system being modelled never exists on the same plane as that being studied methodically.

The point might be clarified by considering Tarski's well-known model-theoretic definition of truth (Tarski [1944]). Is there circularity or regress there? Might it be argued that one needs to know truth before defining it, as Meno would have put it? Of course not, and the same resolution is offered here. Tarski's recursive definition of truth over syntactic construction is based on an appreciation of the properties truth is deemed to have, but that appreciation and the rigorous definition exist on "different planes". So circularity is avoided.

More interesting is the question of infinite regress. Tarski's definition formalises certain specific properties of truth; a regress would obtain only were a complete characterisation sought. So it is with the interpretation required to define an observable. Some property of an undisclosed system is being posited at a certain level of abstraction. An unending sequence of LoAs could possibly obtain were a complete characterisation of a system sought.

It is implicit in the method of abstraction that a GoA is to be chosen that is accurate or "realistic". How, then, is that to be determined without circularity? The answer traditionally offered in mathematics and in science is that it is determined by external adequacy and internal coherence or, in computer jargon, validation (the GoA satisfies its operational goals) and verification (each step in the development of the GoA satisfies the requirements imposed by previous steps). First, the behaviours at a moderated LoA must adequately reflect the phenomena sought by complying with their constraints; if not, then either the definition of the behaviour is wrong or the choice of observables is inappropriate. When the definition of observables must incorporate some "data", the latter behave like constraining affordances and so limit the possible models (see Floridi [2004a] for further details and examples). Second, the condition embodied in the definition of a GoA is a remarkably strong one, and ensures a robust degree of internal coherence between the constituent LoAs. The multiple LoAs of a GoA can be thought of as interlocking like the answers to a multidimensional crossword puzzle. Though such consistency does not guarantee that one's answer to the crossword is the same as the originator's, it drastically limits the number of solutions, making each more likely.

Adequacy/validation and coherence/verification neither entail nor support naive realism. GoAs ultimately construct models of systems. They do not describe, portray, or uncover the intrinsic nature of the systems they analyse. We understand systems derivatively, only insofar as we understand their models. Adequacy and coherence are the most we can hope for.

5. Conclusion

A long time after Gassendi's comment to Descartes, Feynman once remarked that "if we look at a glass of wine closely enough we see the entire universe. […] If our small minds, for some convenience, divide this glass of wine, this universe, into parts – physics, biology, geology, astronomy, psychology, and so on – remember that nature does not know it!"[11] In this paper, I have shown how the analysis of the glass of wine may be conducted at different levels of epistemological abstraction without assuming any corresponding ontological levelism. Nature does not know about LoAs either.

In the course of the paper I have introduced the epistemological method of abstraction and applied it to the study, modelling and analysis of phenomenological and conceptual systems. I have demonstrated its principal features and main advantages. Yet one may object that, by providing a few simple examples and some tailored case-based analyses, the method really predates its applications, which were merely chosen and shaped for their suitability. In fact, it is exactly the opposite: Jeff Sanders and I were forced to develop the method of abstraction when we encountered the problem of defining the nature of agents (natural, human and artificial) in Floridi and Sanders [2004b]. Since then, we have been applying it to some long-standing philosophical problems in different areas. I have used it in computer ethics, to argue in favour of the minimal intrinsic value of informational objects (Floridi [2003]); in epistemology, to prove that the Gettier problem is not solvable (Floridi [2004c]); in the philosophy of mind, to show how an agent provided with a mind may know that she has one and hence answer Dretske's question "how do you know you are not a zombie?" (Floridi [2005a]); in the philosophy of science, to propose and defend an informational approach to structural realism that reconciles forms of ontological and epistemological structural realism (Floridi [2004b]); and in the philosophy of AI, to provide a new model of telepresence (Floridi [2005b]). In each case, the method of abstraction has been shown to provide a flexible and fruitful approach. Clearly, the adoption of the method of abstraction raises interesting questions, such as why certain LoAs, e.g. the so-called

[11] Feynman [1995], the citation is from the Penguin edition, p. 66.

"naive physics" view of the world and the "folk psychology" approach to the mind, appear to be "privileged", or whether artificial life (ALife) can be defined in terms of a GoA. So much work lies ahead.

The method clarifies implicit assumptions, facilitates comparisons, enhances rigour and hence promotes the resolution of possible conceptual confusions. It also provides a detailed and controlled way of comparing analyses and models. Yet, all this should not be confused with some neo-Leibnizian dream of a "calculemus" approach to philosophical problems. Elsewhere (Floridi [forthcoming-a]), I have argued that genuine philosophical problems are intrinsically *open*, that is, they are problems capable of different and possibly irreconcilable solutions, which allow honest, informed and reasonable differences of opinion. The method I have outlined seeks to promote explicit solutions, which facilitate a critical approach and hence empower the interlocutor. It does not herald any sort of conceptual "mechanics".

The method is not a panacea either. I have argued that, for discrete systems, whose observables take on only finitely-many values, the method is indispensable. Nevertheless, its limitations are those of any typed theory. Use of LoAs is effective in precisely those situations where a typed theory would be effective, at least informally. Can a complex system always be approximated more accurately at finer and finer levels of abstraction, or are there systems which simply cannot be studied in this way? I do not know. Perhaps one may argue that the mind or society – to name only two typical examples – are not susceptible to such an approach. In this paper I have made no attempt to resolve this issue.

I have also avoided committing myself to determining whether the method of abstraction may be exported to ontological or methodological contexts. Rather, I have defended a version of epistemological levelism that is perfectly compatible with the criticisms directed at other forms of levelism.

The introduction of LoAs is often an important step prior to mathematical modelling of the phenomenon under consideration. However, even when that further step is not taken, the introduction of LoAs remains a crucial tool in conceptual analysis.

Of course, care must be exercised in type-free systems, where the use of the method may be problematic. Such systems are susceptible to the usual paradoxes and hence to inconsistencies, not only when formalised mathematically but also when considered informally. Examples of such systems arise frequently in philosophy and in artificial intelligence. However, I hope to have shown that, if carefully applied, the method confers remarkable advantages in terms of careful treatment, consistency and clarity.

Acknowledgements

I wish to thank Jeff Sanders, who should really be considered a co-author of this paper, with the exception of any of its potential mistakes; Gian Maria Greco, Jesse F. Hughes, Gianluca Paronitti and Matteo Turilli for their discussions of several previous drafts; Paul Oldfield for his editorial suggestions; Carl Craver for having made his forthcoming research available to me; and finally the anonymous referees of the journal for their constructive feedback.

References

Arbib, M. A. 1989, *The Metaphorical Brain 2 : Neural Networks and Beyond* (New York ; Chichester: Wiley).

Barwise, J., and Etchemendy, J. 1987, *The Liar : An Essay on Truth and Circularity* (New York ; Oxford: Oxford University Press).

Bechtel, W., and Richardson, R. C. 1993, *Discovering Complexity : Decomposition and Localization as Strategies in Scientific Research* (Princeton: Princeton University Press).

Benjamin, P., Erraguntla, M., Delen, D., and Mayer, R. 1998, "Simulation Modeling and Multiple Levels of Abstraction" in *Proceedings of the 1998 Winter Simulation Conference*, edited by D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan (Pistacaway, New Jersey: IEEEPress), 391-398.

Block, N. 1997, "Anti-Reductionism Slaps Back" in *Philosophical Perspectives 11: Mind, Causation, and World*, edited by J. E. Tomberlin (Oxford - New York: Blackwell), 107-133.

Brown, H. C. 1916, "Structural Levels in the Scientist's World", *The Journal of Philosohy, Psychology and Scientific Methods*, 13(13), 337-345.

Craver, C. F. 2004, "A Field Guide to Levels", *Proceedings and Addresses of the American Philosophical Association*, 77(3).

Craver, C. F. forthcoming, *Explaining the Brain: A Mechanist's Approach*).

Davidson, D. 1974, "On the Very Idea of a Conceptual Scheme", *Proceedings and Addresses of the American Philosophical Association*, 47. Reprinted in *Inquiries into Truth and Representation* (Oxford: Clarendon Press, 1984): 183-98. All page numbers to the quotations in the text refer to the reprinted version.

de Roever, W.-P., and Engelhardt, K. 1998, *Data Refinement : Model-Oriented Proof Methods and Their Comparison* (Cambridge: Cambridge University Press).

Dennett, D. C. 1971, "Intentional Systems", *The Journal of Philosophy*, (68), 87-106.

Dennett, D. C. 1987, *The Intentional Stance* (Cambridge, Mass ; London: MIT Press).

Egyed, A., and Medvidovic, N. 2000, "A Formal Approach to Heterogeneous Software Modeling" in *Proceedings of the Third International Conference on the Fundamental Approaches to Software Engineering (Fase 2000, Berlin, Germany, March-April) - Lecture Notes in Computer Science, No. 1783*, edited by Tom Mailbaum (Berlin/Heidelberg: Springer-Verlag),

Feynman, R. P. 1995, *Six Easy Pieces* (Boston, MA.: Addison-Wesley).

Floridi, L. 2003, "On the Intrinsic Value of Information Objects and the Infosphere", *Ethics and Information Technology*, 4(4), 287-304.

Floridi, L. 2004a, "Information" in *The Blackwell Guide to the Philosophy of Computing and Information*, edited by L. Floridi (Oxford - New York: Blackwell), 40-61.

Floridi, L. 2004b, "The Informational Approach to Structural Realism". final draft available as IEG – Research Report 22.11.04, http://www.wolfson.ox.ac.uk/~floridi/pdf/latmoa.pdf

Floridi, L. 2004c, "On the Logical Unsolvability of the Gettier Problem", *Synthese*, 142(1), 61-79.

Floridi, L. 2005a, "Consciousness, Agents and the Knowledge Game", *Minds and Machines*, 15(3-4), 415-444.

Floridi, L. 2005b, "Presence: From Epistemic Failure to Successful Observability", *Presence: Teleoperators and Virtual Environments*, 14(6), 656-667.

Floridi, L. forthcoming-a, "Information Ethics: Its Nature and Scope" in *Moral Philosophy and*

Information Technology, edited by Jeroen van den Hoven and John Weckert (Cambridge: Cambridge University Press),

Floridi, L. forthcoming-b, ""Levels of Abstraction: From Computer Science to Philosophy"", *Journal of Applied Logic*.

Floridi, L., and Sanders, J. W. 2004a, "The Method of Abstraction" in *Yearbook of the Artificial - Nature, Culture and Technology, Models in Contemporary Sciences*, edited by M. Negrotti (Bern: Peter Lang), 177-220.

Floridi, L., and Sanders, J. W. 2004b, "On the Morality of Artificial Agents", *Minds and Machines*, 14(3), 349-379.

Foster, C. L. 1992, *Algorithms, Abstraction and Implementation : Levels of Detail in Cognitive Science* (London: Academic Press).

Gell-Mann, M. 1994, *The Quark and the Jaguar : Adventures in the Simple and the Complex* (London: Little Brown).

Hales, S. D., and Welshon, R. 2000, *Nietzsche's Perspectivism* (Urbana: University of Illinois Press).

Hayes, I., and Flinn, B. 1993, *Specification Case Studies* 2nd ed (New York ; London: Prentice Hall).

Heil, J. 2003, "Levels of Reality", *Ratio*, 16(3), 205-221.

Hoare, C. A. R., and He, J. 1998, *Unifying Theories of Programming* (London: Prentice Hall).

Hughes, P., and Brecht, G. 1976, *Vicious Circles and Infinity : A Panoply of Paradoxes* (London: Cape). Originally published: Garden City, N.Y. : Doubleday, 1975.

Kant, I. 1998, *Critique of Pure Reason* repr. w. corr. (Cambridge: Cambridge University Press). Translated and edited by Paul Guyer, Allen W. Wood.

Kelso, J. A. S. 1995, *Dynamic Patterns : The Self-Organization of Brain and Behavior* (Cambridge, Mass ; London: MIT Press).

Marr, D. 1982, *Vision : A Computational Investigation into the Human Representation and Processing of Visual Information* (San Francisco: W.H. Freeman).

McClamrock, R. 1991, "Marr's Three Levels: A Re-Evaluation", *Minds and Machines*, 1, 185-196.

Mesarovic, M. D., Macko, D., and Takahara, Y. 1970, *Theory of Hierarchical, Multilevel, Systems* (New York: Academic Press).

Nagel, T. 1974, "What Is It Like to Be a Bat?" *The Philosophical Review*, 83(4), 435-450.

Newell, A. 1982, "The Knowledge Level", *Artificial Intelligence*, 18, 87-127.

Newell, A. 1990, *Unified Theories of Cognition* (Cambridge, Mass ; London: Harvard University Press).

Newell, A. 1993, "Reflections on the Knowledge Level", *Artificial Intelligence*, 59, 31-38.

Oppenheim, P., and Putnam, H. 1958, "The Unity of Science as a Working Hypothesis" in *Minnesota Studies in the Philosophy of Science. Concepts, Theories, and the Mind-Body Problem.*, edited by H. Feigl, Michael Scriven, and Grover Maxwell (Minneapolis: University of Minnesota Press), vol. 2, 3-36.

Poli, R. 2001, "The Basic Problem of the Theory of Levels of Reality", *Axiomathes*, 12, 261–283.

Pylyshyn, Z. W. 1984, *Computation and Cognition : Toward a Foundation for Cognitive Science* (Cambridge, Mass: MIT Press).

Robinson, J. 1989, *Vintage Timecharts : The Pedigree and Performance of Fine Wines to the Year 2000* (London: Mitchell Beazley).

Russell, B. 1902, "Letter to Frege" In *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, ed. by J. van Heijenoort (Harvard University Press: Cambridge, MA, 1967), 124-125.

Salthe, S. N. 1985, *Evolving Hierarchical Systems : Their Structure and Representation* (New York: Columbia University Press).

Schaffer, J. 2003, "Is There a Fundamental Level?" *Nous*, 37(3), 498-517.

Simon, H. A. 1969, *The Sciences of the Artificial* 1st ed. (Cambridge, Mass. - London: MIT Press). The text was based on the Karl Taylor Compton lectures, 1968.

Simon, H. A. 1996, *The Sciences of the Artificial* 3rd ed. (Cambridge, Mass. ; London: MIT Press).

Spivey, J. M. 1992, *The Z Notation : A Reference Manual* 2nd ed (New York ; London: Prentice-Hall).

Tarski, A. 1944, "The Semantic Conception of Truth and the Foundations of Semantics", *Philosophy and Phenomenological Research*, 4, 341-376. Reprinted in L. Linsky (ed.) *Semantics and the Philosophy of Language* (Urbana: University of Illinois Press, 1952).

Wimsatt, W. C. 1976, "Reductionism, Levels of Organization and the Mind-Body Problem" in *Consciousness and the Brain*, edited by G. Globus, G. Maxwell, and I. Savodnik (New York: Plenum), 199-267.