ISSN 0104-4443

The foundational crisis of cognitive science: challenging the emergentist challenge

A crise fundacional da ciência cognitiva: desafiando o desafio emergentista

Jean-Michel Roy

Center for the Epistemology of Cognitive Science, University of Lyon, Ecole Normale Supérieure, France, e-mail: jmroy@ext.jussieu.fr

Abstract

The following pages contend that, in spite of its intensive development, contemporary cognitive science has recently entered a phase of fairly acute uncertainty and confusion regarding some of its most essential foundations. They emphasize two aspects of this foundational crisis, specifically vindicating the existence of a crisis of naturalism and of a crisis of representationalism. Like any foundational crisis, this situation constitutes a serious threat to the significance of the empirical achievements of cognitive science. A threat calling for renewed efforts to provide it with secure foundations that can only be obtained through a closer collaboration between empirical and foundational investigations, or, more concretely, between cognitive scientists and philosophers. They also outline a general strategy to address this threat, and illustrate it about one aspect of particular importance of the naturalist crisis, namely the emergentist challenge to the orthodoxy of non reductive functionalism. They argue for the rejection of one version of this emergentist challenge, and they lay out a minimal condition that any other version of emergentism must meet. It is still unclear whether this condition is yet satisfactorily met by some version, and in particular by the sort of emergentism associated with the notion of dynamical system. Clarifying this issue should accordingly be seen as a top priority on the agenda of cognitive philosophy.

Keywords: Cognitive science. Naturalism. Representationalism. Emergentism.

Resumo

As páginas seguintes sustentam que, apesar de seu intenso desenvolvimento, a ciência cognitiva contemporânea recentemente entrou em uma fase de incerteza e confusão razoavelmente intensa em relação a alguns de seus fundamentos mais essenciais. Elas enfatizam dois aspectos desta crise fundacional, especificamente vindicando a existência de uma crise do naturalismo e de uma crise do representacionismo. Como qualquer crise fundacional, esta situação constitui uma séria ameaça à significância das conquistas empíricas da ciência cognitiva. Uma ameaça que exige esforços renovados para muni-la com fundamentos seguros, que somente podem ser obtidos por meio de uma colaboração estreita entre as investigações empíricas e as fundacionais, ou, de forma mais concreta, entre cientistas da cognição e filósofos. Esboçam também uma estratégia geral para tratar desta ameaça, e ilustram um aspecto de particular importância da crise naturalista, a saber, o desafio emergentista à ortodoxia do funcionalismo não reducionista. Elas argumentam em favor da rejeição de uma versão deste desafio emergentista, e expõem a condição mínima que qualquer outra visão do emergentismo deve satisfazer. Ainda é incerto se esta condição tenha sido alcançada de maneira satisfatória por alguma versão, e em particular pelo tipo de emergentismo associado à noção de sistema dinâmico. A elucidação deste problema deveria ser encarada como a prioridade maior no programa de trabalho da filosofia cognitiva.

Palavras-chave: *Ciência cognitiva. Naturalismo. Representacionismo. Emergentismo.*

The idea of a foundational crisis of cognitive science

The contemporary sciences of cognition have been characterized since their inception by a fairly sustained rhythm of evolution, which speaks in favour of their truly scientific nature. In the last two decades, this evolution seems to have taken a particularly dramatic course, with the occurrence of major transformations of several kinds, such as the emergence of new branches of cognitive investigation (e.g. cognitive social neuroscience, affective neuroscience, cognitive ergonomics…), the transformation of older ones (e.g. the appearance of the New Artificial Intelligence...), the sudden blooming of time honoured problems (e.g. the problem of phenomenal consciousness), or the multiplication of alternative frameworks such as Cognitive Neuroscience, the Embodied and Situated Approach (CLARK, 1997), the Enactive Approach (VARELA et al., 1993), the Subjective Approach (LAKOFF, 1987), the Phenomenological Approach (SMITH et al., 2005), the Dynamical Approach (van GELDER, 1999). However, this series of important transformations has not been without creating a certain amount of confusion in the state of cognitive science, for various reasons.

The first one has to do with their nature. With the possible exception of the first of the four just mentioned, most of these transformations do not incarnate a simple process of specification of a well established and clearly delineated architecture of cognitive research, but a deeper one of more or less voluntary revision of the foundations on which cognitive research has so far been based. In other words, most of these transformations amount to foundational challenges, with the effect of obscuring basic problems, concepts and principles of explanation that were thought to be unquestionable, and consequently well understood. Moreover, the shadow cast thereby on the nature of previously accepted foundations naturally extends onto the whole body of results built up with their help.

It is not uncommon these days to see philosophers recoil when hearing about the foundations of a scientific discipline. I nevertheless think that the idea that cognitive science has foundations, and that the specific business of philosophy in the area of cognition is to investigate those foundations, is essentially correct, in addition to being in fact widespread, and that such recoiling is the result of misunderstanding. I will not argue for it here, however, and I will only briefly unfold what I take it to mean. My main concern is to offer a rather precise, but possibly controversial, definition of a thesis that, for widespread as it might be, is unfortunately often left in a state of imprecision that proves to be damaging for the understanding of the function of philosophy in cognitive investigation.

A *foundation* of cognitive science can first simply be characterized as any element introduced as a solution to a *foundational problem*. The category *foundational problem* can in turn be defined as the set of difficulties that are *theoretically prior* to those related to the investigation of specific cognitive phenomena. *Theoretical priority* is different from chronological priority. As well illustrated by the case of mathematics, a science can thus very well develop successfully without having solved its foundations problems: a whole body of mathematical theorems was secured before a decent understanding of the concept of number was reached. However, as also well illustrated by the case of mathematics, until its foundational problems are solved, the results obtained by a science remain uncertain and obscure. As B. Russell wrote at the time of the foundational crisis of mathematics: "although something was true, no two people agreed as to what it was that was true, and if something was known, no one knew what it was that was known" (RUSSELL, [1903] 1980, § 3).

Theoretical priority can be further divided into different kinds, including at least logical priority or priority in the order of truth, epistemological priority or priority in the order of justification, and heuristic priority or priority in the order of discovery. It is also important to underline that theoretical priority is necessarily correlated with theoretical independence: given that solutions to foundational problems command the solutions given to non foundational ones, they are autonomous from them. Various degrees can however be distinguished in such autonomy. Seeing it as absolute gives birth to what might called a foundationalist conception of the foundations of cognitive science, since it largely corresponds to the classical notion of foundationalism, according to which the whole body of human science rests on an independent and absolute knowledge, and which is well illustrated by the traditional understanding of the relations between metaphysics and physics. The modern conception, which I fully share, is different. It only grants a limited form of heuristic and epistemological autonomy to the investigation of foundational problems, and imposes a constraint both of empirical fruitfulness and empirical justification on the solutions brought to them. In this perspective, a solution to a foundational problem is no more than a *foundational hypothesis* that will only be retained if it is capable of generating adequate empirical results.

Finally, foundational problems can be divided into four main categories: 1/ those dealing with the delineation of the domain of investigation of cognitive science, 2/ those dealing with the types of problems that can be legitimately raised about this domain, 3/ those dealing with the methodological and epistemological conception of scientific knowledge that should be used in order to solve these problems, and 4/ those dealing with the basic concepts and principles that should be used in the content of those solutions.

The second reason for the confusing effect of the major transformations of cognitive science is the fact that the foundational challenges they involve are often difficult to grasp, both in their critical and in their constructive dimensions. It is for instance often quite uneasy to understand what exactly antirepresentationalists reject when they propose to eliminate the notion of representation from the framework of cognitive explanation, as well as what they propose to replace it with. The confusion is reinforced by the fact that the foundations being challenged, for well established that they might have been, were often themselves lacking clarity, as some analysts have rightly insisted.[1]

Finally, a third directly correlated reason lies in the absence of clear boundaries between many of these foundational challenges. The Situated and Embodied Approach, for instance, clearly overlaps with the Enactive, the Subjective or the Dynamical ones, the latter one entertaining itself incestuous relations with neo-connectionism. And in spite of a few attempts to disentangle these intricate links, such as the rich and subtle analysis of Andy Clark in *Being There*, their logical geography clearly still stands in need of substantial clarification.

As a result of this confused situation, the general problem of the foundations of cognitive science can be legitimately considered today in a state of irresolution. Two forms of irresolution should however be distinguished. A mild form, which is a normal aspect of scientific investigation in all domains, and in virtue of which foundations are never entirely settled and some foundational issues remain always open, or partially open. And a stronger or deeper form, characterized by the fact that none of the foundational issues is the object of a reasonably widespread consensus, as well as by the correlative presence of radically opposed foundational hypotheses, and leading to strong divergences as to the signification and value of the empirical results already obtained. In my opinion, an important number of the transformations recently

[1] Andy Clark, for instance, very aptly remarks, in discussing some current rejections of the idea that cognitive processes are computational, that such criticism is all the more difficult to circumscribe that the very notion of computation was never made satisfactorily clear by computationalists themselves in the first place (CLARK, 1997, p. 159).

undergone by the contemporary sciences of cognition have made them pass, if not entirely at least to a significant extent, from a situation of mild foundational irresolution to a state of strong foundational irresolution.

This situation is of course a highly unsatisfactory one from a scientific point of view and requires accordingly, urgent clarification. Where do we really stand regarding the foundations of the cognitive enterprise? And what are the appropriate foundations on which it should stand? The basic issues to be addressed in order to overcome this foundational crisis are fairly simple and can be listed as follows:

1) What aspects exactly of the foundations of classical cognitive science are being rejected?
2) Are the interpretations of these foundations on which such rejections are based acceptable?
3) On what grounds exactly are they rejected?
4) Are those grounds valid ones?
5) What is the exact content of the various foundational alternative proposals being made? And in particular, to what extent do they represent real alternatives, both with respect to the foundational hypotheses they put into question and with respect to each other?
6) Finally, to what extent are they themselves well argued for?

These several questions delineate the framework of a systematic critical inquiry into the current foundational situation of cognitive science that represents an indispensable step towards the resolution of the foundational crisis it goes through. Several options as to the results that this inquiry might yield can be envisaged. At one extreme, it might show that this foundational crisis is no more than a tempest in a teapot. At the other extreme, it might as well confirm that contemporary cognitive science is indeed undergoing a thorough process of revision of its initial foundations, which will in the end leave it with a very different face from the one with which it was born.

The naturalist and the representationalist sides of the crisis

Two central aspects of this foundational crisis deserve special attention, given the essential role that the issues they involve have played in the deployment of the contemporary cognitive enterprise: the issue of cognitive naturalism and the issue of representationalism.

The crisis of naturalism

Indeed, the current situation of cognitive science with respect to naturalism clearly conveys a certain sense of disarray. From their very first steps, the contemporary sciences of cognition have massively embraced the perspective of naturalism. The Cognitive Revolution was without any doubt conceived by its protagonists as the triumph of the twentieth century naturalist party. Although it was that of its modern wing over its conservative wing, or in other words, of the rebellious partisans of a non reductionist form of naturalism over the traditional partisans of a reductionist form, who had dominated the scene from Carnap, in 1932, to Herbert Feigl, in 1958. This triumphant non reductive naturalism mainly took the guise of functionalism. And, in spite of the persistence of other views, including that of a neo-reductionist current famously incarnated by Paul and Patricia Churchland, the general feeling was undoubtedly one of an historical breakthrough.

However, the clouds were soon to come and little by little a number of tenacious difficulties accumulated over the head of functionalists, to the point of making it necessary to reopen the entire issue of naturalism in the eyes of a growing number of specialists. And this is why the term ‘crisis’ does not look like too strong a word to describe the current situation. No certitude seems to go reasonably unchallenged anymore, once rather marginal oppositions to the orthodoxy of functionalism – as Ned Block once called it (BLOCK, 1978) – are becoming more center-stage, and researchers seem to be investigating actively again the whole spectrum of possibilities, including the renunciation to naturalism. It is no heresy anymore to envisage publicly that cognitive science should renounce its original alliance with the project of becoming a science of nature, or should at least remarry with such a mild form of naturalism that it becomes disputable whether it still deserves to be called naturalist. Three developments played a prominent role in the dismantling of the functionalist confidence of having finally solved the mind body problem. One is the demonstration of the relative weakness of the main anti-reductionist weapon known as the multiple realizability argument. Epiphenomenalist considerations, and the exclusion problem in particular, showing that functionalist properties are deprived of causal efficacy, represented another major blow. Finally, the famous hard problem has a serious impact on the credibility of functionalism, since the functionalist conception of mental states does not seem to work for qualitative states.

This way of seeing things might sound a bit too dramatic to some, but it is at least shared by a few others. As a matter of fact, the best expression of it is probably due to the philosopher Robert van Gulick, although van Gulick seems to locate the sole source of the difficulty in the so-called hard-problem, that is to say in the problem of naturalizing the phenomenal dimension of consciousness. Indeed, in a paper entitled *Reduction, Emergence and Other recent opinions on the Mind/Body Problem: a Philosophic Overview*, (VAN GULICK, 2001) makes a very similar point, putting it interestingly in Kuhnian terms:

> In Kuhnian terms, physicalism (particularly the sort of functionalistic nonreductive physicalism that has become the mainstream view among philosophers in recent decades) plays the role of normal science, and consciousness (especially the so-called hard problem of explaining how phenomenal consciousness might be just a physical aspect of reality) provides the anomaly that generates the push toward extraordinary theorizing. How the current psycho-physical crisis will be resolved as yet remains unclear, revolutions may or may not be needed (VAN GULICK, 2001, p. 8).

And in order to help resolving the crisis, Van Gulick also agrees on using the same general strategy recommended above. What is needed is thus a rigorous critical analysis of the competing hypotheses in presence, with a view first to determining with accuracy the contents of the various criticisms of the naturalist orthodoxy that have been made, as well as of the various alternative proposals to it that have been offered, and then to assessing the soundness of both of these criticisms and of these proposals. This is the only way to understand whether a naturalist revolution is in the offing, or what we are witnessing is just a moment of temporary naturalist depression.

The crisis of representationalism

The problem of representationalism has many surface similarities with that of naturalism, as well as deep ones. One of these surface similarities is the fact that it has been from the start as much a central foundational problem for contemporary cognitive science as the problem of naturalism. Another is that it has also received a massively positive answer: cognitive science has been from the sixties as dominantly representationalist as naturalist. The question of interest here is whether cognitive representationalism can also be considered as going through a crisis period of the same kind as that currently affecting naturalism. At first sight, the answer is not so clear.

The notion of representation is certainly at the center of many current debates, and some of them can no less certainly be counted as revealing a certain amount of foundational irresolution. As prominent examples of such debates, one could mention, for instance, the dispute over the theoretical versus simulationist character of folk representationalism – what Joseph Perner (1991) has called the representational mind –, the related but different question of whether the explanation of perception and of action requires the introduction of a specific simulationist form of representation, the quarrel over the existence of non conceptual representations, or that of a specifically pragmatic kind of representation. Each one of these debates has a foundational dimension in that it puts ultimately into question the very definition of the most general features of the property of representation, and shows that cognitive science might have initially confused the genus with the species. However, this form of foundational irresolution is more of the mild kind than of the strong kind. It is part of the previously mentioned normal process of revision of foundational hypotheses that usually takes place in scientific investigation. It does not show that the foundations of cognitive science are still fundamentally unresolved, but that they progress. And these debates will at some point abate, as the imagery debate or the opposition between local and distributed representations for instance have.

However, other aspects of the representationalist debate speak more directly to the idea that the problem of representationalism is in a state of deeper and more problematic irresolution, or, more precisely, that it is switching from a state of mild to a state of strong or deep irresolution. In other words still, that our understanding of the problem of representation might have entered a phase of regress more than progress. Five different, although interrelated, kinds of symptoms of such a strong irresolution can be distinguished.

The first one is the development of *substantial challenges* to well established aspects of cognitive representationalism. And chief among them is the development of an anti-representationalist current that questions the very relevance of the notion of representation to the scientific study of cognitive phenomena. Cognitive anti-representationalism has always been around, as the early works of Stephen Stich (1983) or Maturana and Varela (1980) testify, all of them advocating, although in different forms, the elimination of the concept of representation from cognitive science. However, the anti-representationalism seems to have picked up steam over the last twenty years or so, and this from a variety of horizons. It is present, in more or less radical versions, in important work in cognitive neuroscience, such as that of Vittorio Gallese[2] or the later developments of Varela (1993); in artificial intelligence, especially with the development of the New Artificial Intelligence movement under the impulsion of Brooks (1991); in psychology, for instance in the dynamical approach to development of Thelen and Smith (THELEN et al., 1994); in animal psychology, for instance in the criticism that Allen and Berkoff (ALLEN et al., 1995) offered of Fodor's intentional realism in the explanation of animals; and finally, of course, in philosophy, where it comes under a variety of guises in authors such as Dreyfus, Noe and Thompson, van Gelder, Bechtel, Clark, Hendrick Jansen…

The *growing complexity* of the purely philosophical dimension of the representationalist debate can also be seen as a symptom of strong foundational irresolution, in the sense that the multiplication of theoretical positions often betrays an impossibility of reaching a reasonable amount of agreement on essential aspects of the problem of representationalism. The literature on the analysability of the notion of qualia in representational terms is a good case in point, or almost any aspect of the analysis of the nature of representation (its inner structure, its source of determination, its content…).

A correlate of these first two symptoms is a *growing inconclusiveness* of certain aspects of the representationalist debate. Many of its core questions have been there for almost fifty years now and show little signs of receding. On the contrary, since even some of the most consensual ones, such as that of the relevance of representation, are being reopened.

A fourth symptom is the fact that the problem of representationalism is intrinsically connected with many other foundational issues, and that the recent transformations affecting those other issues have *potential implications* of great importance for it. As a result, it can hardly be considered as satisfactorily solved until these implications have been thoroughly explored. The various facets of the problem of representationalism have for instance been mainly investigated in the perspective of cognitivism, where the psychological level of inquiry is considered as heuristically independent from the implementation

[2] For instance, GALLESE, 2001.

(neurobiological) level. But the development of the cognitive neuroscience approach is increasingly challenging this heuristic autonomy and, as a consequence, transforming the answers to many aspects of the problem of representation, as the whole current of the philosophy of cognitive neuroscience has begun to show.

Finally, in spite of its richness, it is arguable that the representationalist debate is still suffering from a few important lacunae, such as the neglect of the phenomenality of representation or the complexity of the relations between representation and intentionality, and that, as long as these lacunae are not addressed, the problem of representation cannot be considered as satisfactorily solved, because they also have substantial implications for its resolution.

Taking into consideration these various aspects of the irresolution of the problem of representationalism, it seems to me that its similarity with the present state of the problem of naturalism extends to the idea that there also is a representationalist crisis.

A crisis that requires the same type of general strategy of resolution recommended above. Indeed, what is needed in order to overcome it is, in the first place, a rigorous theoretical definition of the problem of representationalism. Even though this problem is more intuitively graspable than the problem of naturalism, I am afraid that the representationalist debate frequently suffers, however, from an insufficiently rigorous understanding of it. And what is also needed is, in the second place, a detailed critical analysis of its present state with a view to sorting out with precision its points of mild and deep irresolution, to determining with exactitude for each of them what the hypotheses in presence are, what these hypotheses really claim, how they actually relate one to each other, and which one, if any, should be favoured. It is, for instance, of first importance to elaborate a clear concept of what a non-representationalist explanation of cognition might be, and also to get a clear picture of the multifaceted growing anti-representationalist current, of what its various versions really are, how much they differ one from each other, of how well taken are the criticisms of representation they put forward… None of these questions has at this moment a clear and straightforward answer. When one looks carefully at the anti-representationalist positions currently defended, it is manifest that they are not without misunderstandings and vagueness about the representationalism that they reject, that their proposals are not deprived of confusion and inconsistency, so that in the end one wonders whether they are not after all variants of representationalism themselves. And indeed, I think that the concept of representation has often been, and can be, defined in such a simple and fundamental way that the most basic aspect of the problem of representationalism is not: should a theory of cognition be representationalist or not? But: how could a theory of cognition not be representationalist?

Can emergentism resolve the naturalist crisis?

Defining and assessing neo-classical emergentism

The naturalist side of the foundational crisis is in many ways its most important aspect, since the commitment to cognitive naturalism of cognitive science commands the totality of its other foundational hypotheses. And in the first place that of representationalism. From its perspective, the property of representation can only be rehabilitated, against the radical rejection it suffered from the behaviorist paradigm, under the condition that it can be naturalized. This is the reason why I would like to develop one aspect of the resolution strategy recommended above as it applies to the problem of naturalism, focussing specifically on the pretension of emergentism to offer a valuable alternative to non reductive functionalism.

Emergentism is indeed, with neo-psychoneural reductionism, one of the two main challengers to the functionalist orthodoxy. One reason for putting the emphasis on emergentism is that it is clearly gaining momentum and, after a long period of quasi-absence in the mind-body problem – with, as usual, a few noticeable exceptions such as those of Searle (1983) or Bunge (BUNGE et al., 1990) –, making sort of a comeback, a comeback which is in fact not limited to the epistemology of the cognitive sciences. A few data will suffice to substantiate the claim. The 1992 landmark publication of a collection of essays on the subject by Beckerman, Flhor, e Kim, (1992) under the title *Emergence or reduction: essays on the prospect of Nonreductive naturalism* (BECKERMAN et al., 1992) launched the trend. 1997 saw the publication of an influential issue of the French *Intellectica* Journal on the theme of "Emergence and Explanation" (CASATI, 1997). Several international meetings followed in the early 2000s about different facets of emergentism and various collective volumes, at times related with them, have been published since.The second reason is that this revival originates in various schools of thinking and that it is important to confront the various forms that what might be called neo-emergentism assumes as a result of this diversity of sources. A third one is the fact that, despite its growing importance in the naturalist debate, this neo-emergentism has not yet received sufficient critical attention. Finally, a last one is the obvious conceptual closeness of the notion of emergentism with that of non reductionism that makes it a priority to determine whether emergentism can provide a more adequate form of non reductive naturalism than functionalism.

Accordingly, one crucial and still rather unexplored problem about the naturalist crisis of cognitive science is whether emergentism can indeed bring it an adequate solution. A problem that requires a thorough critical assessment of the current emergentist challenge organized around the following interrogations:

1) what are the various types of emergentism currently explored, and how do they relate one to each other?
2) what naturalist claims do they make? To what extent, in particular, can they be assimilated indeed with a non reductionist form of naturalism? And to the extent that they can, wherein lies their specificity?
3) what criticisms do they address to other naturalist doctrines?
4) is any of them capable of overcoming the difficulties at the source of the crisis of naturalism?

There are several possible reasons for which emergentism might turn out to be of no help. A notion with a bad reputation, emergentism might for instance prove to be too difficult to define with sufficient precision. Or it might be that it is in fact an ontologically neutral notion, and that it is more challenging than it looks to turn it into a significant naturalist one. Finally, it might also be that the type of naturalism that it can provide does not differ substantially enough from others, and in particular from non reductive functionalism.

The way I propose to carry out the critical inquiry which I see as indispensable to reach a solid decision among these various options comprises two steps. The first one consists in focussing on one of the different types of emergentism currently at work in the cognitive science area, and in offering a definition of its characteristics as well as an assessment of its naturalist virtues. The results obtained are then used to clarify by differentiation what are, if any, the other forms of emergentism in competition, and whether they do any better in solving the difficulties that led to distrust functionalism. It is the conclusions of the first one of these two steps that, capitalizing on previous work, I wish to expose in the following paragraphs, where I will defend four successive claims:

1) a salient component of the emergentist challenge is a type of emergentism that deserve to be labelled neo-classical, inasmuch as it is the legacy of the original XIX century British notion of emergentism;
2) this neo-classical emergentism does not seem capable to offer a way out of the exclusion problem faced by non reductive functionalism;
3) accordingly, the challenge that the emergentist challenge itself is facing is that of offering a form of emergentism that does provide, *at least*, a solution to the exclusion problem;
4) in this perspective, a priority issue on the agenda of whoever is concerned with solving the naturalist crisis is to figure out whether any other current emergentist challenger does provide one, and in particular so called dynamical emergentism.

The problem of cognitive naturalism

The naturalist dimension of this neo-classical emergentism, as well as its specificity with regards to non emergentist naturalist doctrines, can only be adequately grasped on the background of a precise understanding of the problem it is supposed to solve, and the way these other doctrines propose to solve it.

It is not uncommon to see the notion of naturalism being accused of elusiveness. I think there is little ground to this accusation and that we know very well what this notion means, even if providing a fully satisfactory elucidation of it is, as always, a bit challenging. Indeed, naturalism is just a specific form of monism, and monism can in turn be characterized at its most general level in terms of properties and as the thesis that a scientific theory (in general or relative to such or such domain) should recognize as relevant to its goal only one category of properties. Accordingly, cognitive naturalism can itself be defined as the thesis that all properties deemed as relevant in an adequate scientific investigation of cognitive phenomena should be natural ones. In other words, cognitive naturalism asserts that a scientific theory of cognition should use one category of property only, namely the category of natural property.

If we schematise the scientific knowledge of a domain D as follows:

$$\begin{tikzcd} \boxed{\mathbf{T}} \arrow[r] & \boxed{\begin{matrix} \mathrm{x}\,[\boldsymbol{P} = \{\mathrm{P}_1, \mathrm{P}_2, \ldots, \mathrm{P}_n\}] \\ \\ \mathbf{D} \end{matrix}} \end{tikzcd} \tag{1}$$

cognitive naturalism can be defined as the simple thesis that for **D** = cognition, $\boldsymbol{P} = \boldsymbol{P}_N$.

In fact, this definition should be refined a bit with a distinction between the properties that characterize thereoy T as a scientific operation, or epistemological properties, and the properties that characterize the content of T, and might be labelled ontological properties, in the general sense that they are properties attributed to elements of its domain. In other words, $P = P_E + P_O$. Accordingly, one should distinguish between epistemological and ontological naturalism, that is to say between $P_E = P_N$ and $P_O = P_N$. And in order that a theory of cognition be a fully naturalist one, it should obviously subscribe to both forms of naturalism. That is to say that it should only make use of natural properties in the characterization of its object, and in the characterization of itself as a theorizing activity. However, the current debate around naturalism remains mostly restricted to the ontological side of the problem, and I will follow this bad example, since our main interest is here of a critical nature.

If cognitive naturalism is the thesis that a theory of cognition should recognize as ontologically relevant properties no other properties than natural ones, *the problem of naturalizing a theory of cognition* is just that of finding a way to conform with this requirement. And it is arguable that it can only be met if, for any ontologically relevant property P, P is either immediately recognized as a natural property, or shown to be derivable from properties immediately categorized as natural by T. Hence any naturalist theory of cognition must determine a subset of *basic* or *primitive* natural properties, and devise a derivation procedure to obtain from it another subset of *derived natural properties*. The first task requires to first define the concept of a basic natural property, and then to determine whether any property satisfies it. The second task is the core of the naturalization problem and consists in finding a principle of transformation of apparently non natural properties into natural ones.

Any such principle of naturalization *proprio sensu* must respect three basic conditions:

a) *the attribution constraint*: the property P to which it applies must be considered as belonging to an x that is a natural entity, that is to say to an entity with natural (primitive or previously derived) properties;
b) *the ontological constraint*: P must be recognized to belong to x in virtue of the natural properties that characterize x as a natural entity, and be thereby ontologically dependent on them, in the general sense of having the instantiation of these natural properties as necessary and sufficient condition for its own instantiation;
c) *the explanatory constraint*: This ontological dependency must also be rationally explained; in other words, a principle of naturalization must provide a rational principle of explanation of the fact that those natural properties are necessary and sufficient conditions of instantiation of property P.[3]

The general structure of an ontologically naturalist theory of cognition can accordingly be represented with the following schema, where, for sake of simplification, P_{NP} designates a primitive natural property, P_{ND} a derived natural property, the down arrow the ontological relation of dependency, and the up arrow the explanatory relation:

$$\begin{tikzcd}
 & P_{ND} \arrow[dl] \arrow[d, shift right, very thick] \\
x & P_{NP} \arrow[l] \arrow[u, Rightarrow, shift right]
\end{tikzcd}$$

(2)

[3] This third requirement explains why treating the ontological dependency as a brute fundamental fact, as property dualism – of a certain kind at least – does, is a borderline case of naturalism. On the one hand, it can be read as a failure to fulfil the explanatory constraint. But on the other hand, it can be read as fulfilling it, although pointing to the limits of explainability. To adopt this second reading implies that one should accept the idea that there is no other way of explaining a certain number of things than by being rationally entitled to say: this is just how things are.

This schema captures the three fundamental elements of the idea of naturalization as a derivation of natural properties, namely that an apparently non natural property is turned into a natural one when 1/ it is attributed to an entity with properties already categorized as natural, 2/ it is supposed to belong to it in virtue of these primitive or previously derived natural properties, and 3/ a rational explanation of the fact that it belongs to it in virtue of its natural properties is offered. And this schema can be easily recast simply in terms of mental and neurobiological properties to make it more consonant with the immediate way of understanding the problem of cognitive naturalism:

$$\begin{tikzcd}
 & M \arrow[dl] \arrow[d, thick, shift right] \\
x & N_{eur} \arrow[l] \arrow[u, Rightarrow, shift right]
\end{tikzcd} \tag{3}$$

In conclusion, the *problem of cognitive naturalism* can now be simply defined as that of deciding whether or not a theory of cognition should conform with the requirement of using only natural properties, and if it does, as that of finding a way to fulfil adequately the two tasks just laid out for all properties seen as relevant to the investigation of cognitive phenomena, especially for mental ones.

It is important to underline that according to this definition, any theory that tries to conform to this double requirement deserves the title of cognitive naturalism, independently of the results of its efforts. A distinction is thereby drawn between theories that do not embrace a naturalist perspective on cognition, and those that fail to do so. The definition is also broad enough to include naturalist theories both of an internalist kind, that make all properties dependent on properties internal to the cognitive system, and of an externalist kind, that make them also, at least partially, dependent on natural properties external to the cognitive system.

Once defined, the problem of naturalism offers interesting similarities with the problem of representationalism. As a matter fact, the two problems have a similar structure: both are about the relevance of certain properties for obtaining an adequate scientific theory of cognition. The problem of naturalism, as we have seen, is fundamentally that of determining whether *all cognitively* relevant properties should be natural ones, and if so, how such a requirement could be respected. In a similar fashion, the problem of representationalism is fundamentally that of determining whether relevant properties – and especially ontologically relevant ones – should *include* the property of representation, as well as related ones. Accordingly, the basic difference between the two problems lies only in the content of the relevant property under discussion, and in the type of relevance of this property – exclusive versus non-exclusive. And an appropriate definition of representationalism is precisely one that specifies in detail the nature of the representational properties considered to be relevant as well as the type of relevance that they have.

On the basis of this fundamental characterization of the problem of representationalism, it is possible to unfold somehow a priori what its full-blown structure must be. There are different ways of doing it, but all of them will include a same number of inescapable issues.

First comes the question of the *eliminability* of the property of representation. A positive answer to this question essentially terminates the investigation. But a negative one, or at least a partially negative one, to the effect that some cognitive states and processes at least should be treated as representational by an adequate theory of cognition, immediately leads to the further question of the *nature* of the property of representation and of related properties (representational properties) that have been recognized as relevant. Two key aspects of this question are the analysis of the general structures of representation (the reference/content distinction, the logical properties of representational idiom…) and the variety of formats of representation. From here, one can jump directly to the question of the *ontological value* of the representational properties so defined, and in particular to the debate between representational realism and irrealism: are representational properties really determinations of cognitive systems or merely instruments of prediction of cognitive phenomena? A correlate of this question is the problem of the *causal efficacy* of representational properties, and in particular of their content. A further issue is the problem of the *naturalization* of the property of *representation*, and especially the naturalization of its causal efficacy. Then comes the question of the *determination* of the various representational properties, and therefore debates such as the one between internalism and externalism, or between a descriptive versus a non descriptive view of reference-fixing. Finally, one should also mention at least the problem of the relations between linguistic representation and mental representation, and between representation and phenomenal consciousness.

The main solutions to the problem of cognitive naturalism

In the perspective opened by the above definition, the standard solutions to the problem of cognitive naturalism can be seen essentially as variations, firstly, on the choice of properties at the base of the schema (behavioural properties for behaviorism, physical properties for physicalism…) and, secondly, on the choice of a principle of derivation. From this last point of view, the standard way of classifying naturalist solutions is to draw the main line of division between reductionism on the one hand, and non reductionism on the other one, non reductionism being itself mainly subdivided into token physicalism, functionalism and emergentism.

It is impossible to go into the details of this standard interpretation of reading the logical geography of modern cognitive naturalism. The issue is indeed quite complex. Functionalism has for instance both a reductionist version (LEWIS, 1972; KIM, 1998) and a non reductionist one; and token physicalism is understood in at least two different ways, that make it both compatible or incompatible with functionalism. I will therefore limit my analysis to what is essential for clearly apprehending the location, within this general picture, of classical emergentism, to which I will devote more attention.

The hallmark of *cognitive reductionism*, at least in its central and classical form illustrated by Logical Behaviorism or the Central State Identity theory of the 60s, is to accept an identity relation between types of mental properties and non disjunctive types of natural properties, in such manner that every token of a mental property of type say M_1 is identical with a token of a non disjunctive natural property of type say N_1. Being identical with natural properties, be they behavioural or neurobiological, mental properties are necessarily ontologically dependent on them, and the fact that these natural properties are necessary and sufficient conditions of their instantiation is also trivially explained. Accordingly, cognitive reductionism can be schematised as follows:

$$\begin{tikzcd}
 & M \arrow[dl] \arrow[d, line width=2pt] \\
x & N \arrow[l] \arrow[u, Rightarrow]
\end{tikzcd}$$

(4)

$\boldsymbol{\longrightarrow}$ = type identity

Non reductionism, on the contrary, refuses such a type identity relation of this kind.

In its most simple form, namely that of *token physicalism*, a type of mental property is conceived as being identical with a disjunctive type of natural property, so that in each one of its instantiations or tokenings a mental property of type say M_1 is identical with a natural property N, although not always of the same type (it might be of type N_1 or N_2 or N_3…).

As for *functionalism*, at least according to one dominant reading of it, it identifies a type of mental property with a type of functional property that simply depends on variable types of natural properties, without being either identical to any of them or to their disjunction. This dependency relation is variously characterized as a relation of implementation, realization or supervenience, the basic definition of supervenience being that a set of properties such as properties M supervenes on another set of properties such as properties N when the instantiation of properties M is fixed once the instantiation of properties N is fixed, but not vice-versa (so that two natural entities cannot have different mental properties if they have identical natural properties, but can have different natural properties if they have identical mental properties). When functional properties are more specifically understood in a causal way, the fact that mental properties belong to an entity in virtue of its natural properties is explained by its assimilation with the supposedly unproblematic fact that the causal properties of a natural entity are ontologically dependent on its natural properties. Functionalism thus corresponds to the following version of the above schema:

$$\begin{tikzcd}
 & M = F \arrow[dl] \arrow[d, line width=2pt, shift left=-2] \\
x & N_1 \vee N_2 \vee N_3 \arrow[l] \arrow[u, Rightarrow, shift right=-2]
\end{tikzcd} \qquad (5)$$

$\boldsymbol{\longrightarrow}$ = realization/implementation/supervenience

It should be observed that, if reductionism is simply defined as a form of type-identification of mental properties with natural properties, however, functionalism can be seen as a form of reductionism, since a type of causal property of a natural property is a type of natural property, although a second order one. This is the reason why it is in fact more appropriate to define reductionism in a more restrictive way, as the identification of types of mental properties with non disjunctive types of *first order* properties, and consequently to define non reductionism as any solution to the problem of naturalism that refuses such identification.

The solution of neo-classical emergentism

In reviewing these fundamental distinctions of the naturalist debate, my main concern is to bring to light the specificity of emergentism considered as a non reductionist naturalist doctrine. As already mentioned, this non reductionist view of the relations between emergentism and the naturalist issue is essentially the fact of one current of emergentism, and it is on this current that I wish to concentrate. It is certainly not the only one to be taken into consideration, but it is the one that seems to have implicitly dominated the emergentist challenge. Getting a clear view of what it claims to achieve in this respect, and of what it has really achieved, is consequently an important step towards reaching an adequate assessment of this challenge.

The current of emergentism under consideration might, as already indicated, be dubbed neo-classical emergentism, since it is closely connected with the tradition of reflection on emergence inherited from the philosophy of scientific knowledge of John Stuart Mill, and it has played a dominant role in the development of emergentist views. It therefore includes what Brian McLaughlin has dubbed 'British Emergentism'[4] as its main source and element. Achim Stephan (1992) has proposed to distinguish four main phases in the evolution of classical emergentism so characterized, and it is important to go over some aspects of this development in order to clarify the relation of classical emergentism with the problem of naturalism.

The starting point: Mill's theory

Following A. Stephan, the first initial phase is the introduction of the concept and term of emergence in the second half of the XIX century in Mill's

[4] McLAUGHLIN, 1992.

System of Logic in 1843,[5] Bain's *Logic* (1870),[6] and finally Lewes' *Problems of Life and Mind* (1875),[7] who is responsible for the actual philosophical invention of the term. The second phase is the search for an alternative to both mechanism and vitalism in the analysis of the relations between physics and biology that took place in the 1920s, and was mainly carried through in the investigations of Samuel Alexander (Space, 1920 *Time and deity*),[8] Lloyd Morgan (1923) *Emergent Evolution*, and Broad (1925) *The Mind and its place in Nature*. The protracted debate that resulted from these investigations and extended up to the definitions of emergence by Hempel ([1948] 1965) and Nagel (1961) constitutes a third phase, with a substantial number of contributions, including those of Pepper (1926), Stace (1939), Henle (1942), Pap (1951). Finally, the progressive reinvestment of the notion into the more restricted mind-body debate starting in the late 70's with Mario Bunge (1980) and Popper et al. (1977) or Roger Sperry (1980) opens a fourth phase, clearly still under way, and might we add, picking up steam.

Let us have a closer look at the first two phases, given their theoretically crucial character.

Mill introduces the notion that will be later called emergence in the context of a discussion of causality, and in order to distinguish two types of complex causation, understood as a causal process involving several causes, or, as Mill puts it, a "composition" of causes. The difference between the two types lies in the difference between the resulting effect of this composition of causes and the effect that would have resulted from the isolated actions of the various causes involved in the composition. In other words, the difference between the two types of complex effects lies in their respective relations to what might be called the related isolated effects.

In one case, the complex process of causation is of the same nature as the isolated processes of causation, and the difference between the complex effect and the related isolated ones is therefore purely quantitative. It is just a combination of these effects, proportional to the combination of the causes. Such is the case, according to Mill, for complex causation in the field of mechanics:

[5] MILL, [1843] 1973.

[6] BAIN, 1870.

[7] LEWES, 1875.

[8] ALEXANDER, 1920.

> If a body is propelled in two directions by two forces, one tending to drive it to the north, and the other to the east, it is caused to move in a given time exactly as far in both directions as the two forces would separately have carried it; and it is left precisely where it would have arrived if it had been acted upon first by one of the two forces, and afterwards by the other. (MILL, [1843]1973, p. 210).

In the other case, the complex process of causation is not similar in nature with the isolated ones, and the difference between the complex effect and the related isolated ones is therefore qualitative. In other words, when several relations of causality are combined, they produce an effect of a different type, or content, or nature from the ones they would produce in isolation. Mill writes: "In certain cases, single causal relations change the nature of their effects when acting in a compound manner, since the effect is no more present in the complex effect as a part of it". Such is in his eyes the case with chemical causation, although he does not analyze it rightly, since his analysis insists on the difference between the nature of the causes and the nature of the complex effect, which is in fact irrelevant, instead of insisting on the difference in nature between the effect obtained in the chemical reaction and the related effects that would have been obtained if the causes had operated in isolation.[9] The heart of the difference between the two cases is thus that, in the first case complex causation does not give rise to an unprecedented relation of causality, while it does in the second case. It is ultimately a difference between "laws which work together without alteration, and laws which when called upon to work cease and give place to others" (MILL, [1843] 1973, p. 211). The first type of complex causation is variously called by Mill homopathic or homogeneous causation, and obeys a principle of composition of causes asserting that "the joint effect of several causes is identical with the sum of their separate effects". And he calls the second type heteropathic or heterogeneous causation. And it is for this very notion that Lewes later coined coin the term 'emergence'.

[9] Mill writes: "The chemical combination of two substances produces, as is well known, a third substance with properties entirely different from those either of the two substances separately, or of both of them taken together" (MILL, [1843]1973, p. 210).

Mill's theory calls for a number of remarks

1) The notion of emergence – although the term is not there yet – is intrinsically linked to that of causality: it characterizes a specific form of causation relation.
2) It is also intrinsically linked to the notion of complexity, in the sense that emergent causation is a *specific* form of complex causation.
   However, it is not intrinsically linked to it in the sense that it would be a characteristic of a mereological part-whole relation. Indeed, the complex relation of causality and the complex effect are not said to be emergent with respect to their components, but with respect to isolated relations of causality and to their effects. The whole point is precisely that such relations and effects disappear in the case of emergent causation.
3) It is intrinsically linked, on the other hand, with the notion of naturalism, in the sense that emergent causation is a type of natural causality; the difference between emergent and non emergent causality "is one of the fundamental distinctions in nature" (MILL, [1843] 1973, p. 211).
   However, it not intrinsically linked with the problem of naturalism as previously defined, that is to say with the problem of finding a way to respect the constraint of making all scientifically relevant properties natural ones.
4) The difference between emergent causality and non emergent causality is an ontological one that has an epistemological correlate, namely the deducibility of the complex effect in the first case, and the non deducibility of the complex effect in the second one. In the case of homopathic causation, Mill writes:

> We can compute the effects of all combinations of causes…from the laws which we know to govern those cases when acting separately. Not so in the phenomena which are the peculiar object of the science of chemistry…. we are not, at least in the present state of our knowledge, able to foresee what result will follow from any new combination, until we have tried it by specific experiment (MILL, [1843]1973, p. 210).

However, emergence is not defined in terms of non deductibility or non predictability, which are introduced essentially as epistemological correlates of emergence, but in terms of non compositionality: an emergent effect is not the sum of the related isolated effects; it is not the whole that results from the aggregation of the isolated effects. What is emergent for Mill is a non compositional complex effect, and it is emergent inasmuch as it is non compositional.

5) Finally, nature seems to be organized in different domains linked by relations of emergent causation, that delineate as many scientific domains of investigation, and where the ultimate causal responsibility seems to fall on physical entities, although this idea of a layered structure of the universe is not fully developed.

The resulting general picture of emergentism

From Mill's initial step, classical emergentism seems to have evolved, according to several of its analysts, in the direction of 1) a dissociation of emergence from causality, transforming it into a structural relation more than a causal one; 2) the establishment of an intrinsic association with the idea of a part-whole relation, so that what is emergent is, in contradistinction to Mill's view, the properties of a whole with respect to the properties of its parts; 3) the establishment of an intrinsic association with the problem of naturalism and the search for a principle of naturalization; 4) the characterization of the content of the notion of emergence in terms of non deducibility, and hence non reducibility; 5) and, finally, a stronger affirmation of a layered view of reality.

In his 1992 article "'Downward causation' and Emergence", J. Kim proposed for instance to summarize this general conception of emergentism born out of the tradition originated by Mill with the three following theses:

1) [Ultimate physicalist ontology] There are basic, nonemergent entities and properties, and these are material entities and their fundamental particles.
2) [Property emergence] When aggregates of basic entities attain a certain level of structural complexity ("relatedness"), genuinely novel properties emerge to characterize these structured aggregates. Moreover, these emergent properties emerge *only* when appropriate 'basal' conditions are present.

3) [The irreducibility of Emergents] Emergent properties are ‘novel’ in that they are not reductively explainable in terms of the conditions out of which they emerge.

The paradigmatic illustration of Broad’s theory

And indeed the work of D. C. Broad, in which the second phase of the evolution of classical emergentism culminates, seems to fit pretty well this general picture. According to the central view it puts forward, reality is a mereological layered structure, divided into a hierarchy of types of aggregates of basic material entities, and some of these types of aggregates have characteristic properties which are emergent, in the sense that they are irreducible to the properties of their components.

In such a view, emergence is essentially a structural property of dependency, since it characterizes the way how the characteristics of an aggregate depend on those of its components. It does have also a dynamical dimension, since the aggregate is supposed to result from a process of aggregation, and this process of aggregation is a causal one. But this dimension does not seem anymore to be essential, in the sense that, were the aggregate not the result of a causal process, it would still be characterized as emergent with respect to its components.

For the same reason, emergence is also a mereological part-whole relation, a relation between the macro-properties and the micro-properties of an entity.

In addition, it is explicitly offered as a solution to the problem of naturalism, inasmuch as it is introduced in order to avoid the dualist implications of the vitalist response to the difficulties encountered by the reductionist analysis of the relations between biology and physics.

Finally, emergence is assimilated with irreducibility, and not with non-compositionality; and irreducibility, in turn, is assimilated with non-deducibility.

In order to clarify these various points let us examine more closely Broad’s definition of emergence.

Although the notion of emergentism is introduced by Broad in the context of the explanation of the characteristic properties of living bodies, and more specifically, of the rejection of a vitalist form of explanation, this problem is in his eyes a special case of the broader one of explaining the characteristic properties of complex entities, complex entities being understood as structured aggregates. He distinguishes several possible ways of explaining these characteristic properties. Emergentism is one of them, and, as in Mill, is to be opposed to mechanism.

Emergentism and mechanism are both seen as being themselves different from a form of explanation that rejects the notion that the properties of a whole are in any way determined and explainable by the properties of its parts, and also from so called component explanations – or, better said, 'special component' explanations –, which claim that it is necessary to postulate special components in order to account for the specific properties of wholes. Vitalism belongs, according to Broad, to this category in that it claims that a special component, called entelechy, is necessary in order to account for the characteristic properties of living bodies.

However, emergentism and mechanism also differ from each other in the following way. A mechanistic account, according to Broad, accepts that the determination of the properties of a whole by the properties of its parts obeys general principles, so that the specific properties that obtain in a whole W, resulting from the instantiation of a relation R among components A, B and C, can be deduced from the knowledge of the nature of components A, B and C, and of these general principles. Emergentism, on the contrary, denies such possibility in a number of cases, thereby also rejecting the notion that general principles regulate in these cases the determination of the specific properties of W by its components. In such cases, of which chemical entities are again taken as the paradigmatic example, the determination is considered as whole specific and depending uniquely on R. Broad does not exactly phrase the difference in those terms, speaking instead of the possibility or impossibility of deducing the specific properties of a whole W, whose internal composition is expressed as R(A,B,C) from "the most complete knowledge of the properties of A,B and C *in isolation* or *in other wholes which are not of the form R*(*A*,*B*,*C*)" (BECKERMANN, 1992). But if the specific properties of R(A,B,C) could be deduced from the properties of A,B,C and those they have in other wholes R'(A,B,C), R''(A,B,C), it would mean precisely that there are general principles governing the determination of wholes by their component parts, and that they apply to R(A,B,C). And asserting, on the contrary, that the specific properties of the structure R(A,B,C) depend uniquely on R is denying that they result from the application of general principles of determination of macro-properties by micro-properties, to use a more modern language.

Broad explicitly confirms this interpretation with his further distinction between general laws and "ultimate and unique" laws. The first ones govern the relation of the specific properties of different types of wholes with those of their components, while the second ones apply to a single type of whole. He writes for instance about the specific properties of the chemical substance silver-chloride:

> […] it would be useless to study chemical compounds in general and to compare their properties with those of their elements in the hope of discovering a general law of composition by which the properties of any chemical compound could be foretold when the properties of its separate elements were known; so far as we know, there is no general law of this kind… the properties of silver-chloride with those of silver and chlorine and with the structure of the compound is, so far as we know, an unique and ultimate law (BECKERMANN, 1992, p. 106).

Ansgar Beckerman aptly reformulates Broad's definition of emergence in the following way:

> Let S be a system having the microstructure [C_1…..C_n; R]; then F is an emergent property of S iff:
> a) There is a law to the effect that all systems with this microstructure have F, but
> b) F cannot, even in theory, be deduced from the basic properties of the components C_1…C_n and a general theory of components of this kind which contains no unique and ultimate laws which apply only to systems which have the same microstructure as S. (BECKERMANN, 1992, p.106).

Such a formulation, however, leaves aside the opposition to reductionism that came to be seen as a prominent feature of classical emergentism, and that is present indeed in Broad's doctrine since, again, it was designed as a middle way between vitalism and reductionism, even though, from a terminological point of view, it is explicitly more opposed to mechanism than to reductionism. It is easy, however, to show that the notion of non-deducibility is very closely connected with that of non-reducibility, since, according to the prevailing definition of reduction at least, which corresponds to the so-called theory model of reduction well brought out by early logical positivism, and whose classical formulation is due, on the one side, to Hempel ([1948] 1965) and Nagel (1961) and on the other side to Oppenheim, Kameny and Putnam (OPPENHEIM et al., 1958), a reduction is essentially a *deductive* operation, since it amounts to integrating one deductive system – a theory – into another. In this perspective, what Broad calls the non deducibility of the specific properties of an emergent type of whole can be read as the impossibility, for the laws that make use of these properties, to be deduced from those that make use of the properties of the components of the whole.

However, although Broad certainly believes in such an impossibility, this assimilation is wrong since it does not take into account the fact that, in Broad, the notion of non reducibility fundamentally applies to properties. What is irreducible is first and foremost the specific properties of the whole, and in a secondary way only, the laws in which they enter. So, if one is to think in terms of the theory model of reduction, Broad's notion of non reducibility is more appropriately assimilated with a failure to fulfil what Nagel labelled the condition of connectability. The condition of connectability states that in order to deductively integrate a theory into another, their properties have to be connected by principles that make them homogeneous enough to make the deductive operation possible. One way to fulfil this condition is by identifying the properties of the reduciendum theory with logical constructions of properties of the reduciens theory. The laws of thermodynamics, according to this analysis, can for instance be deduced from those of statistical mechanics through the identification of the property of temperature with that of mean kinetic energy of molecules. Now such identification is indeed what Broad in essence denies, since his notion of non deducibility is a rejection of the possible identification of the specific properties of certain types of wholes with the properties resulting from the nature of their components and general principles of combination. Emergent properties are non resultant properties, and they cannot be identified with resultant properties because resultant properties are characteristically not Disponível em such cases. So, one could say that Broad's emergentism is opposed to reductionism in the theory model sense of the term in that it sustains that certain wholes cannot fulfil the requirement of connectability understood as a requirement of identifiability.

One objection, however, can be opposed to this analysis. It has been voiced by several commentators, including Beckerman and Kim in their 1992 precious collection of essays on emergentism (BECKERMANN, 1992; KIM, 1992). Its basic idea is that, even though in emergent wholes resultant properties are not available for identification with the specific properties of the whole, these specific properties are still coextensional with a certain microstructure, and could therefore be identified with it. For instance, even though W_1 with microstructure R[A,B,C] has, among its specific properties no resultant property F_1 coextensional with an emergent property F_2, F_2 is still coextensional with R[A,B,C], and could therefore be identified with it by Broad. The question raised by this objection is whether the fact that Broad's emergentism refuses such identification mean that it cannot be assimilated with a non-reductionism in the sense theory model sense of reductionism, and if it cannot, to determine the real concept of reduction that it opposes. In other words, the sort of coextension required by reductionism in the theory model sense of the term is available to Broad. And the problem is therefore whether Broad disregards them because his emergentism does not oppose reductionism in the model theory sense, but in some other sense to be determined.

Such is the opinion of Beckerman and Kim, although but they diverge as to the nature of that alternative form of reductionism. Beckerman, for instance, sees it as a sort of anticipation of the modified version of the theory model of reduction due mostly to C. Hooker, P. Churchland and more recently J. Bickle (1998). In this modified version, reduction remains an operation of deductive intregration, but connecting principles, especially of the coextensional kind, play no essential role. Another possibility is to see it as an anticipation of a conception of reductionism that would reject more radically the theory model, such as the one vindicated for instance by Bechtel and Richardson who write in "Emergent phenomena and Complex Systems":

> Within the biological and non physiological sciences, there is something else that counts as reduction, albeit not theory reduction. A reductionist in biology or psychology is someone who seeks to explain the key phenomena that have been recognized at one level of organization in nature and that have more commonly been identified by or proposed as a result of inquiries pursued in another discipline (BECHTEL et al., 1992).

The stumbling block of epiphenomenalism

The best way to assess the neo-classical emergentist solution is to examine how much it can convincingly pretend to solve the various difficulties imputed to non reductive functionalism. And in one respect at least, I do not think that its claim to do so fares well. As a matter of fact, neo-classical emergentism looks unable to overcome the so-called ‘exclusion problem’, to the effect that non reductive functionalism makes psychological properties causally inefficacious, because their causal efficacy is absorbed by the neurobiological properties acting as the necessary and sufficient conditions of their instantiation. I have argued the point in detail elsewhere and will only briefly summarize my argument here Roy (2004).

J. Kim, who is probably the first to have pinpointed the difficulty, gave it its classical formulation (KIM, 1993). This formulation rests on many presuppositions and in particular on a disputable nomological account of causality. Assuming the validity of such an account, the exclusion problem can be stated about the central case of psychophysical causation, where one mental psychological property – say Psy 1 – is declared the efficient cause of a physical property – say Phy2 – of a behavioural kind, in the simple following terms. If the fact that Psy1 causes Phy2 means nothing else, in virtue of the nomological interpretation of causality, that the fact that an instantiation of Psy1 is nomologically followed by an instantiation of Phy2, and if, in virtue of the principle of non reductive naturalism, Psy1 cannot be instantiated without a physical property – say Phy1 – of a neurobiological kind being instantiated, the nomological relation between Psy1 and Phy2 is inseparable from a nomological relation between Phy1 and Phy2. In other words, the psychophysical relation of causation Psy1 ➞ Phy2 is inseparable from a physical relation of causation Phy1 ➞ Phy2. Furthermore, it is arguable that the first one is redundant with respect to the second one, and should therefore be abandoned in its favour, given that the second is more fundamental in virtue of the ontological dependency of the instantiation of Psy1 on the instantiation of Phy1.

My first claim is that, even though the exclusion problem has been raised in the context of a criticism of functionalism, it is rooted in features that functionalism shares with other forms of non reductive naturalism and that constitute its very defining characteristics. More specifically, it is bound with the notion of abstract and irreducible property, that is to say, with the notion of a property whose instantiation is ontologically dependent upon the instantiation of another property – and consequently only entertains an abstract form of independence from it –, but that cannot nevertheless be reduced to it. As a consequence, any solution to the problem of cognitive naturalism that makes use of the notion of an abstract and irreducible property is liable to the exclusion problem. And my further claim is that neo-classical emergentism does make use of it, and therefore represents no progress over functionalism on the exclusion problem. An emergent property, as conceived by neo-classical emergentism, is indeed a property whose occurrence depends on the occurrence of a number of other properties seen as its conditions of emergence, and which is at the time irreducible to them.

Conclusion: the challenge to the emergentist challenge

If the above argument is correct, emergentism cannot claim to provide a better solution than functionalism to the problem of cognitive naturalism without offering a way out of the exclusion problem, and consequently, without also finding a way to overcome the apparent limitations of the notion of an abstract and irreducible property. In this lies a challenge for the emergentist challenge itself to functionalism. A challenge that represents a philosophical priority, given its stakes for cognitive naturalism at large, and for the cognitive science enterprise in particular.

This challenge has two main aspects. One is to try to devise an emergentist solution to the exclusion problem and to examine in the first place whether this problem can in principle be solved by emergentist means, or whether its source runs so deep that the very notion of emergentism is doomed to failure. And the other is to extend the critical examination of emergentist doctrines beyond neo-classical emergentism, in order to determine whether some other emergentist candidate might already offer a more convincing proposition in this respect.

Dynamical emergentism is one of these candidates, and it should be privileged in such an examination for two reasons at least. One is its growing importance in cognitive science, as in science in general, and the other is the conceptual vagueness that still seems to affect nevertheless the concept of emergence with which it operates. And I would like to end in this regard with a personal worry that seems to make the challenge to emergentism even more challenging.

Indeed, it has somehow passed unnoticed that dynamical emergentism is in fact reintroducing a crucial element of Millian emergentism, namely the fact that an emerging property is an effect and emergence consequently a causal relation. The point is for instance made quite clear in the dynamical emergentism associated with F. Varela's cognitive enactivism. In *Radical embodiment: neural dynamics and consciousness*, Varela and E. Thompson (2001) claim for instance that consciousness should be seen as a property emerging from large scale synchronization (through phase locking) of neuronal activity, and this process of emergence is unambiguously assimilated to one of "upward-causation".

From the point of view of the exclusion problem, this reactivation of a causal perspective on emergentism might be not disadvantageous. As a matter of fact, Jackson e Pettit (1990) rightly observe that the problem does not arise when the two causes involved are sequentially organized, because what appears to be a competition of causes is replaced with a cooperation of causes within a causal chain. The instantiation of Phy1 becomes, for instance, the first element of a causal process whose end result is the instantiation of Phy 2, and the instantiation of Psy 1 becomes an intermediary step in this process.

It is to be feared, however, that there is a heavy price to pay for this advantage because the exclusion problem seems to be circumvented at the expense of what Kim himself called the 'pairing problem' of causality.[10] Indeed, dynamical emergentism so construed solves the psychophysical causation problem by just accepting it as a basic fact: it is a fact of nature that, under certain conditions, a qualitatively novel property such as a mental one is the effect of physical ones. And the truth is that such a position looks perfectly consistent with a nomological conception of causality. If causality is nothing more than nomological constant conjunction, why would psychophysical constant conjunction be more problematic than physico-physical conjunction? However obvious, the point remained largely unaddressed in the mind body problem literature and Kim deserves credit for confronting it in relation with the interactionism of Cartesian substance dualism: if causality is nothing more than constant conjunction, the causal interaction between a material and a spiritual substance looks no more problematic than a constant conjunction between two mental substances. Kim argues that the feeling that there nevertheless is in this case a special difficulty is well grounded. And he sees the heart of this difficulty in the fact that the nomological conception of causality requires "a shared space-like coordinate system in which the objects are located, a scheme that individuates objects by their locations in the scheme",[11] and that physical space seems to be the only such coordinate system available. The problem being therefore that the spiritual substance cannot be located in physical space. In other words, certain homogeneity of nature between two objects is required for a nomological conjunction to take place, and spatiality

[10] KIM, 2006.

[11] KIM, 2006, p. 71.

looks like the only available homogeneous determination. But this lack of homogeneity between conjoined elements seems in fact to affect all the same the causal emergence of a mental property from physical ones within a supposedly unique substance. And it is therefore arguable that dynamical emergentism ignores that emerging properties of the sort that can be qualified as mental are somehow more radically novel than emerging properties of the sort that qualify as physical, and that it does not do philosophically better than Cartesian interactionism by accepting psychophysical causation as a *datum naturae* of an emergentist kind.

References

ALEXANDER, S. **Space, time and deity**. London: Macmillan, 1920.

ALLEN, C.; BERKOFF, M. Cognitive ethology and the intentionality of animal behavior. **Mind & Language**, v. 10, p. 313-328, 1995.

BAIN, A. **Logic**. London: Longmans, 1870.

BECHTEL, W.; RICHARDSON, R. Emergent phenomena and complex systems. In: BECKERMANN, A.; FLOHR, H.; KIM, J. (Ed.). **Emergence or reduction?** essays on the prospect of nonreductive physicalism. Berlin: Walter de Gruyter, 1992. p. 257-288.

BECKERMANN, A. The historical facet of emergence. In: BECKERMANN, A.; FLOHR, H.; KIM, J. (Ed.). **Emergence or reduction?** essays on the prospect of nonreductive physicalism. Berlin: Walter de Gruyter, 1992. p. 102.

BECKERMANN, A.; FLOHR, H.; KIM, J. **Emergence or reduction?** essays on the prospect of nonreductive physicalism. Berlin: Walter de Gruyter, 1992.

BICKLE, J. **Psychoneural reduction**. Cambridge, MA: MIT Press, 1998.

BLOCK, N. Troubles with functionalism. **Readings in the Philosophy of Psychology**, v. 2, p. 128-134, 1978.

BROAD, C. D. **The mind and its place in Nature**. New York: Harcout, Brace & Company, 1925.

BROOKS, R. Intelligence without representation. **Artificial Intelligence**, v. 47, p. 67-90, 1991.

BUNGE, M.; ARDILA, R. **Philosophy of psychology**. New York: Springer-Verlag, 1990.

BUNGE, M. A. **The mind-body problem**: a psychobiological approach. Oxford: Pergamon Press, 1980.

CASATI, R. Emergence and explanation. In: CASATI, R. (Ed.). Intellectica. 1997.

CLARK, A. **Being there**: putting brain, body and world together. Cambridge, MA: MIT Press, 1997.

GALLESE, V. The ‘shared manifold’ hypothesis: from mirror neurons to empathy. **Journal of Consciousness Studies**, v. 8, n. 5/7, p. 33-50, 2001.

HEMPEL, C.; OPPENHEIM, P. Studies in the logic of explanation. In: HEMPEL, C. (Ed.). **Aspects of scientific explanation and other essays in the philosophy of science**. New York: The Free Press, [1948] 1965. p. 245-296.

HENLE, P. The status of emergence. **Journal of Philosophy**, v. 39, n. 18, p. 486-493, 1942.

JACKSON, F.; PETTIT, P. Program explanation: a general perspective. **Analysi**s, v. 50, n. 2, p. 107-117, 1990.

KIM, J. Downward causation. In: BECKERMANN, A.; FLOHR, H.; KIM, J. (Ed.). **Emergence or reduction?** essays on the prospect of nonreductive physicalism. Berlin: Walter de Gruyter, 1992. p. 49-93.

______. The non reductivist’s trouble with mental causation. In: KIM, J. **Supervenience and mind**: selected philosophical essays. New York: Cambridge University Press, 1993. p. 336-357.

______. **Mind in a physical world**: an essay on the mind-body problem and mental causation. Cambridge, MA: MIT Press, 1998.

______. **Philosophy of mind**. Boulder: Westview Press, 2006.

LAKOFF, G. **Women, fire, and dangerous things**: what categories reveal about the mind. Chicago: University of Chicago Press, 1987.

LEWES, G. H. **Problems of life and mind**. London: Kegan Paul, 1875.

LEWIS, D. Psychophysical and theoretical identifications. **Australasian Journal of Philosophy**, v. 50, p. 249-258, 1972.

McLAUGHLIN, B. P. The rise and fall of british emergentism. In: BECKERMANN, A.; FLOHR, H.; KIM, J. (Ed.). **Emergence or reduction**: essays on the prospects of nonreductive physicalism. Paris: Walter de Gruyter, 1992. p. 49-93.

MILL, J. S. **A system of logic**. Toronto: University of Toronto Press, [1843] 1973.

MORGAN, L. **Emergent evolution**. London: Williams & Norgate, 1923.

NAGEL, E. **The structure of science**: problems in the logic of scientific explanation. New York: Harcourt Brace & World, 1961.

OPPENHEIM, P.; PUTNAM, H. Unity of science as a working hypothesis. In: FEIGL, S.; SCRIVEN, M.; MAXWELL, G. (Ed.). **Minnesota studies in the philosophy of science**. Minneapolis: University of Minnesota Press, 1958. p. 3-27.

PAP, A. The concept of absolute emergence. **British Journal for the Philosophy of Science**, v. 2, p. 302-311, 1951.

PEPPER, S. Emergence. **Journal of Philosophy**, v. 23, p. 241-245, 1926.

PERNER, J. **Understanding the representational mind**. Cambridge, MA: MIT Press, 1991.

POPPER, K.; ECCLES, J. C. **The self and the brain**. Berlin: Springer, 1977.

ROY, J.-M. Conception émergentiste du mental et explication causale. In: DURAND, M.-J. (Ed.). **Des lois de la pensée au constructivisme**: conceptions et modélisations de l'acte de connaître. Paris: Intellectica, 2004.

RUSSELL, B. **The principles of mathematics**. London: Norton, [1903] 1980.

SEARLE, J. R. **Intentionality, an essay in the philosophy of mind**. New York: Cambridge University Press, 1983.

SMITH, D.; THOMASSON, A. **Phenomenology and philosophy of mind**. Oxford: Oxford University Press, 2005.

SPERRY, R.W. Mind-Brain interaction. **Neuroscience**, v. 6, p. 109-113, 1980.

STACE, W. Novelty, indeterminism and emergence. **Philosophical Review**, v. 48, n. 3, p. 296-310, 1939.

STEPHAN, A. Emergence: a systematic view of its historical facets. In: BECKERMANN, A.; FLOHR, H.; KIM, J. (Ed.). **Emergence or reduction?** essays on the prospect of nonreductive physicalism. Berlin: Walter de Gruyter, 1992. p. 119-139.

STICH, S. P. **From folk psychology to cognitive science**: the case against belief. Cambridge, MA: MIT Press, 1983.

THELEN, E.; SMITH, L. B. **A dynamic systems approach to the development of cognition and action**. Cambridge, MA: MIT Press, 1994.

THOMPSON, E. Radical embodiment: neural dynamics and consciousness. **Trends in Cognitive Science**, v. 5, n. 10, p. 418-425, 2001.

Van GELDER, T. Defending the dynamical hypothesis. In: TSCHACHER, W.; DAUWALDER, J.-P. (Ed.). **Dynamics, synergetics, autonomous agents**: nonlinear systems approaches to cognitive psychology and cognitive science. Singapore: World Scientific, 1999. p. 241-255.

Van GULICK, R. Reduction, emergence and other recent opinions on the mind/body problem: a philosophic overview. **Journal of Consciousness**, v. 8, n. 9/10, p. 1-34, 2001.

VARELA, F. **Autonomie et connaissance**: essai sur le vivant. Paris: Seuil, 1980.

VARELA, F.; THOMPSON, E.; ROSCH, E. **The embodied mind**. Cambridge, MA: MIT Press, 1993.

Received: 02/10/2010
Recebido: 10/02/2010

Approved: 05/04/2010
Aprovado: 04/05/2010