Logic, Epistemology, and the Unity of Science 24

Olga Pombo
Juan Manuel Torres
John Symons
Shahid Rahman *Editors*

Special Sciences and the Unity of Science

Springer

SPECIAL SCIENCES AND THE UNITY OF SCIENCE

LOGIC, EPISTEMOLOGY, AND THE UNITY OF SCIENCE

VOLUME 24

Logic, Epistemology, and the Unity of Science aims to reconsider the question of the unity of science in light of recent developments in logic. At present, no single logical, semantical or methodological framework dominates the philosophy of science. However, the editors of this series believe that formal techniques like, for example, independence friendly logic, dialogical logics, multimodal logics, game theoretic semantics and linear logics, have the potential to cast new light on basic issues in the discussion of the unity of science.

This series provides a venue where philosophers and logicians can apply specific technical insights to fundamental philosophical problems. While the series is open to a wide variety of perspectives, including the study and analysis of argumentation and the critical discussion of the relationship between logic and the philosophy of science, the aim is to provide an integrated picture of the scientific enterprise in all its diversity.

For further volumes:
http://www.springer.com/series/6936

Special Sciences and the Unity of Science

Edited by

Olga Pombo
Universidade de Lisboa, Portugal

Juan Manuel Torres
Universidad Nacional de Cuyo, Argentina

John Symons
University of Texas at El Paso, USA

Shahid Rahman
Université de Lille, France

Springer

Editors
Olga Pombo
Faculdade de Ciências
Universidade de Lisboa
Campo Grande
Lisboa, Portugal

Juan Manuel Torres
Facultad de Filosofía y Letras
Universidad Nacional de Cuyo
Campus 'Parque Gral. San Martin'
Mendoza, Argentina

John Symons
Department of Philosophy
University of Texas at El Paso
Worrell Hall 306
El Paso, Texas, USA

Shahid Rahman
UFR Philosophie/UMR-STL: 8163
Université de Lille
Villeneuve d'Asque, France

ISBN 978-94-007-2029-9 e-ISBN 978-94-007-2030-5
DOI 10.1007/978-94-007-2030-5
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2012930991

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Chapter 1
Introduction

Olga Pombo, Juan Manuel Torres, John Symons, and Shahid Rahman

As it was expressed in the *Introduction* to our first volume, in order to clarify the discussions surrounding *unity of science*, a sharp distinction between science unity and science unification should be drawn. Whereas the former demands tasks for the identification of common factors among the diverse disciplines – mainly objects and methods – the latter involves the determination of those formal and material conditions that make the connections between theories really possible. It is of paramount importance to take into account that both conditions should allow for an effective connection between theories.

A quasi trivial consequence of the distinction between unity and unification is that any work of unification contributes to the thesis of unity; though factors of unity will not be able to offer arguments in favour of unification. Some contributions presented in this volume – specific case studies – propose sound arguments in favour of the unification thesis, whereas others point out different factors for unity. In addition, other articles provide analogies, approaches and models which are

O. Pombo (✉)
Faculdade de Ciências, Center for Philosophy of Science, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
e-mail: opombo@fc.ul.pt

J.M. Torres
Facultad de Filosofía y Letras, Universidad Nacional de Cuyo, Campus ‘Parque Gral. San Martin’, 5500 Mendoza, Argentina
e-mail: jmvtorres@yahoo.com.ar

J. Symons
Department of Philosophy, University of Texas at El Paso, Worrell Hall 306,
El Paso, TX 79968, USA
e-mail: jsymons@utep.edu

S. Rahman
UFR Philosophie/UMR-STL:8163, Université de Lille, B.P. 60149,
59653 Villeneuve-d’Asq, France
e-mail: shahid.rahman@univlille3.fr

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology, and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_1,

valuable for rethinking the unity of science. Such works are based in new theories and disciplines some of which were unknown at the time of Neurath and Carnaps' programmes.

In the spirit of the hypothesis of science unification, three papers – by J. R. Croca, E. Beggs et al. and A. Perea – present innovative views. Croca does so by proposing the *principle of eurhythme* as suitable to give unification to the basic laws of Classic, Quantum and Gravity Physics. On their side, Beggs and co-workers point out the relationships between computability and physical systems and present a proposal even stronger than the mere association by asserting that computers exist in nature. Finally, Perera and Sokolić's contribution is done by analyzing the nature of liquids and associated liquids. These authors show how the view they suggest can overcome the double vision of matter that Physics has up today, i.e., continuous and discontinuous. Contributions by N. Gontier and F. Carrapiço describe the importance of symbiosis as a fundamental mechanism of organic evolution and, therefore, as an axis of any evolutionary theory. Paraphrasing the famous dictum by T. Dobzhanski 'nothing makes sense in Biology outside the evolutionary perspective', Gontier and Carrapiço's conclusions seem to encourage the idea that in evolutionary theory nothing makes sense outside the symbiotic perspective. Gontier goes even further by proposing symbiogenesis as a source of metaphors for sciences and an alternative for a universal selectionist account of evolution.

The tension between unity and specialization is focused on by L. Faucher. Based on S. Mitchell's typology of inter-theoretic relationships, the author defends pluralism as a middle way between global inter-theoretic unification and local forms of unification. He offers an analysis of the connections between neuroscience and the cognitive psychology of racial prejudice as an example of local unification, enriching, at the same time, the original proposal by Mitchell.

Is evolutionary biology a part of computer science? This provocative question is addressed by P. Hunemann, who tries to clarify the relationships between computer science and evolutionary biology, connections powered by the development of artificial life studies. Computer studies would provide pure possible processes for Biology to explore the dynamics of real life, despite the fact that actual processes deviate from these idealized models. In this way evolutionary computational models can act as useful null hypothesis for biological change. In this chapter, the reader will find not only elements in favour of the unity of science but also a strong thesis for unification: the evolutionary theory would be, according to the author, a part of computer science.

From an innovative perspective, in fact a model taken from thermodynamics, V. Neves encourages the construction of science as a theoretical open system for breaking not only the dualist view of science in terms of unity/disunity but also trends of integration/disintegration. By analyzing an attempt of universalizing Economics as the centre of social sciences, the author demonstrates the convenience of building theories as a system capable of communication between themselves. Along the same line, K. Laurent argues the possibility of a priori establishing rational principles of a unified science. To do so, she uses the well known Lakatosian notion of scientific research programme, something that allows exploration of the internal problems of disciplines, which are continuously generate by the contradictory hypotheses we find inside science. She exemplified the convenience of adopting the Lakatosian notion by examining, in the realm of Economics, the case of neoclassic view *vis a vis* the heterodox one. At the same time, her adopted approach allows to clarify the characteristic concepts of *pluridisciplinarity, interdisciplinarity and transdisciplinarity.*

Certainly, the notions of *theory* and *science* are not equivalent. However, science in its supreme stage is framed by theories and an intricate web connecting them. So it is not a surprise that the concept of theory, together with the analysis of the relationships among theories, constitutes a topic of great importance for the challenge of unification. In this sense, the structuralist view of theories – an epistemological proposal formulate by W. Stegmüller, J. Sneed and others – should be pointed out as one philosophical school most compromised with the creation of a new notion of theory especially apt for that role. Such a notion is useful for tasks of integration of knowledge that comes from different disciplines. The article by Lorenzano, focused on Freud's theory of hysteria, develops a detailed account of the structuralist view and shows why its notion of theory is appropriate for the integration role, particularly by the introduction of Sneed's highly innovative idea of partial models in science.

In the above mentioned articles and others, the readers will find many arguments and indisputable examples in favour of the unity of science. It is true that there are scenarios and many examples of the disunity as well. But as argued in the *Introduction* of the first volume, one has to take into account that science is a dynamic process in which the assimilation of new phenomena, perspective and hypotheses and their subsequent organization in the scientific *corpus* takes place just slowly. Therefore, a permanent landscape of apparent disunity is an unavoidable consequence of the scientific gradual integration process. Some thinkers prefer to label this recurrent circumstance 'crisis', but they seem to ignore its temporary nature. Only a retrospective view on the practical results of scientific enterprise and on the science itself, grant us a plain testimony of the tendency to unity which characterizes all human knowledge.

Chapter 2
Pragmatic Continuities in Empirical Science: Some Examples from the History of Astronomy

María de la Concepción Caamaño Alegre

Abstract The present work constitutes an attempt at identifying those pragmatic aspects of empirical science that contribute to its continuity. The source of scientific continuity has remained unclear after the incommensurability thesis dramatically challenged the truth-cumulative view of science (which was developed within the traditional view of empirical science as a corpus of theories). The first part of the paper provides a clarification of the pragmatic approach adopted here, in many points close to Nicholas Rescher's methodological pragmatism. Although, as opposed to it, truth and usefulness are kept totally separate on the present account, being the latter just related to empirical soundness instead. The second part of the paper offers an application of such account to some historical examples, which will illustrate the pragmatic continuities displayed by the development of astronomy, from the Babylonian period to Copernicus.

Keywords Ancient astronomy • Pragmatism • Scientific continuity • Theory change

The purpose of this work is to emphasize the importance of those pragmatic aspects of empirical science that contribute to its continuity. The term 'pragmatic' will here be used in two different (although related) senses. First in the sense of 'experiential', following the original pragmatic tradition started by C.S. Peirce, and second in

---

M. de la Concepción Caamaño Alegre (✉)
Stanford University, Ventura Hall, Rm 29 220 Panama Street MC 4101, Stanford,
CA 94305-4101, USA
e-mail: mcaamano@stanford.edu

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology,
and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_2,

the sense of 'epistemically useful', relying on more recent developments of this tradition made by N. Rescher and G. Schurz, among others.[1]

After the more traditional analysis of empirical science as a corpus of theories has faced serious difficulties in keeping the common intuition that scientific development is, to a certain extent, cumulative in nature, and the incommensurability thesis dramatically challenged the truth-cumulative view of science, the source of scientific continuity remains unclear. In what follows I argue that valid procedures, rather than true beliefs, are the essential source of continuity and cumulative development in science. Or, accordingly, that approximately valid procedures, rather than approximately true beliefs, constitute such source of continuity. This idea entails a shift in emphasis from theories to practices, and, thus, from truth evaluation to validity evaluation. The main reason for this change of emphasis is the much higher independence of experimental procedures with respect to theories than the other way around. This is made evident in both the historical survival of useful procedures through theory change, and the theories' dependence on such procedures to be successfully developed. This view, of course, is not new. In Philosophy of Science, authors like Nicholas Rescher or Ian Hacking have made similar claims.[2] Within the post-Kuhnian tradition, both Rescher's methodological pragmatism and Hacking's influential new experimentalism meant a first step towards a recognition of experimental practice as the key aspect determining scientific development. Nevertheless, the specification of the pragmatic features of science that this recognition would demand is still tentative and disperse in the current literature in the Philosophy of Science.[3] This work is meant, therefore, as an extension of some of these author's ideas, more in particular, as an attempt to identify some of the pragmatic features that show especially constant in science. To this end, I will examine the pragmatic continuities displayed by the development of astronomy, from the Babylonian period to Copernicus. I would like to show how, in astronomy, similar kinds of observations, as well as alike computational and predictive procedures, made possible the transition from a flat and static earth at the center of the universe to a round, triple motion earth turning around the sun.[4] The examples are meant to show how useful formal and experimental procedures typically survive through theory change, and develop in a cumulative fashion.

---

[1] For a clarification on the main different senses in which the term 'pragmatism' has historically been used in the current literature on the subject, see Susan Haack's introduction of Haack (2006), pp. 15–67, pp. 16–19.

[2] The views of these authors are presented in Rescher (1977), and Hacking (1983).

[3] It is noteworthy that even in Susan Haack's recent monograph on pragmatism no contemporary discussion of the pragmatic dimension of science is included among the works within the new pragmatism (Cfr. Haack 2006).

[4] This idea is emphatically underlined by one of the most acknowledged figures in history of astronomy, O.E. Neugebauer. In his article "Exact Science in Antiquity" (1941), we find: "The method and even the general mental attitude of the work of Copernicus is much more closely related to that of Ptolemy, a millennium and a half before, than to the methods and concepts of Newton, a century and a half later" (in Neugebauer 1983b, p. 23).

The paper consists of two parts. In the first, a pragmatic claim about the cumulative character of valid (or approximately valid) procedures is made and distinguished from traditional pragmatic claims, concerning the nature of true beliefs. In the second, some examples taken from the history of astronomy are provided as historical evidence showing how pragmatic continuities typically occur in science.

2.1 The Pragmatic Claim About Valid Procedures

It must be clarified, first, that no particular pragmatic approach is neither proposed nor assumed here. Nonetheless the view adopted here is characterized by contrasting it with the traditional pragmatic approaches.

2.1.1 The Traditional Pragmatic Claim Concerning Belief Justification

The basic assumption shared by the different pragmatic accounts of scientific development is that success in attaining certain goals justifies our beliefs about the world.[5] This assumption is extremely general and ambiguous, and yet it has the following clear implications:

1. Success implies effective performance, which directly refers to practice. Justification of beliefs is, thus, dependent on practice, rather than on any *a priori* system of justification. This entails a rejection of any *a priori* conception of belief justification, based on some *a priori* conception of knowledge or knowledge capacities.[6]
2. Success in attaining certain goals constitutes the criterion of belief justification. Since something is successful in so far as it is useful to attain certain goals, usefulness is the ultimate criterion of belief justification.

Letting the above implications aside, most points involved in the pragmatic assumption need clarification. The most urgent one concerns the specification of the kinds of goals pursued in scientific practice. This is a prior question, since the kind of success pursued can only be determined once the goals have been

---

[5] See Gerhard Schurz’s article (Schurz 1998, p. 39).

[6] As P.P. Wiener points out, this idea leads to a fallibilist theory of knowledge. Success implies effective performance, which directly refers to practice. Justification of beliefs is, thus, dependent on practice, rather than on any *a priori* system of justification. Pragmatism entails therefore a rejection of any *a priori* conception of belief justification, based on some *a priori* conception of knowledge or knowledge capacities (Cfr. Wiener 1973–74).

determined, and only after the former has been done it is possible to specify, in turn, the kind of belief justification proposed. In his article "Kinds of Pragmatisms and Pragmatic Components of Knowledge", Gerhard Schurz offers some clarification of these issues. He distinguishes between "pragmatism to the right", or k(nowledge)-internal pragmatism (represented, among others, by C.S. Peirce, C.I. Lewis, the late H. Putnam, N. Rescher and G. Schurz), and "pragmatism to the left", or k(nowledge)external pragmatism (represented, e. g., by W. James, F.C.S. Schiller, R. Rorty and S.P. Stich). According to the former, scientific practice would be guided by epistemic purposes (like observation, prediction, explanation, control, innovation, acquisition of true beliefs, and so on), while, according to the latter, it would be guided by non-epistemic purposes (such as survival, adaptation, technological development, wealth creation, social welfare, and so on).

2.1.2 The Revision of the Traditional Pragmatic Claim

Pragmatists have traditionally emphasized that inquiry is a kind of practice, and refused any *a priori* justification of knowledge, pointing out the justificatory role of practice. The spirit of the work presented here conforms to this general pragmatic trend.[7] But it differs from the traditional pragmatic views in that no pragmatic criterion of belief justification is endorsed. The pragmatic assumption mentioned before will thus be revised in a substantial way. This revision primarily concerns two aspects: the object of justification, and the character of the justification. As to the former, practice, instead of beliefs, will be considered as that what is properly subject to justification in terms of usefulness.

That usefulness is essentially involved in practice justification, and not so in belief justification, has been convincingly argued for by N. Rescher (Rescher 1977). In distinguishing between thesis-pragmatism and method-pragmatism, he maintains that only the second is correct. Thesis-pragmatism would be undermined by two separate facts: that there exist useless truths and useful falsities, and that the usefulness of beliefs it is not amenable to treatment according to logical

---

[7] In "Pragmatism in Physics" (1998), Patrick Suppes argues that the focus on practice as opposed to foundations is clearly recognizable in modern physics, and discusses the different, coexisting interpretations of quantum mechanics relying on the same experimental data (Cfr. Schurz 1998, pp. 246–251). He also draws attention to other pragmatic aspects of science for which the present work provides some further historical evidence. First, the fact that "there is much broad agreement by both theoretical and experimental physicists on the truth or falsity of many kinds of observations made with or without refined instrumentation" (p. 237). Second, "the pragmatic way in which physicists (. . .), can use observations, computations and fragmentary theoretical models from many different viewpoints over many different centuries and take from the past work just that which is relevant and relatively sound" (p. 238).

principles.[8] That is, from the conjunction of two useful beliefs is not possible to derive the usefulness of any of them. A very popular line of pragmatism (initiated by W. James and J. Dewey) has traditionally proposed usefulness as a criterion of truth and, therefore, equated belief justification on the basis of usefulness with belief justification on the basis of truthfulness. The view held here is that, for the reasons given by Rescher, truth cannot be determined in terms of usefulness. The continuities of empirical science are not explained here on the basis of truth (neither understood pragmatically, nor metaphysically), but just on the basis of s(cience)-internal usefulness. The latter would be broader than k-internal usefulness, allowing for practical usefulness, which is not directly epistemic.

It could be argued that this change of focus, from beliefs to practices, renders the pragmatic assumption trivial and insufficient to ground any pragmatic view of science. There is, nevertheless, another factor that preserves the pragmatic relevance of the assumption, namely, the asymmetric relation between practice justification and belief justification. As the historical cases make evident, belief justification relies more heavily in practice justification than the other way around. And this seems especially clear if we examine the relation between both relative to scientific theories where they have been used. Usefulness of practices involved in experimentation has shown highly independent of theoretical beliefs. Measurement, computation, control and manipulation of entities or processes employed in scientific inquiry can be successful regardless of the truth or falsity of the theoretical beliefs implied by the theory for whose verification or refutation those practices are being employed.[9] Practice justification comes first in science, and develops in a highly cumulative way with great independence of scientific beliefs related to the theories in which certain practice is applied.

Practice acceptability, as opposed to belief acceptability, depends always on validity rather than on truth. The validity of a practice or procedure, in turn, depends on its effectiveness to attain the expected result. Some procedures are clearly extensions or refinements of previous ones. Unlike theories, valid procedures are often still considered acceptable even when replaced by others. When one theory is replaced by another, the empirical generalizations characteristic of the former are considered false, whereas when a procedure is replaced by another, in many cases, the former is not considered invalid (but just less precise, complete or convenient). In fact, valid procedures that have been replaced in the current scientific practice can still be effectively used later in new scientific contexts, as will be shown is Sect. 2.2.

So much as to the first revision of the pragmatic assumption, the one concerning the object of pragmatic justification. Let's turn now to the second revision, that related to the specific character of the justification, that is, to the kind of usefulness involved in the pragmatic justification (of practice). The pragmatic approach endorsed here is in tune with the k-internal-pragmatism, and, therefore,

---

[8] See *ibid.*, chapter 4.

[9] Numerous examples illustrating this point can be found in Hacking (1983), chapter 9.

places emphasis on epistemic usefulness. Whereas, from the k(nowledge)external pragmatism, no restrictions related to usefulness are introduced. Epistemic goals rather seem to be questioned, and certain social goals favored instead. On the other hand, the reasons for the epistemic restriction on usefulness are not explicitly stated by the advocates of the k(nowledge)-internal pragmatism. That is probably because some basic assumptions about the nature of empirical knowledge are taken for granted. Like, for example, the idea that empirical knowledge provides information about natural phenomena, and depends on the mensurability and predictability of certain phenomena. The idea that scientific practice has certain essential features that ultimately determine the requirements for its justification may be at the bottom of this lack of explanation for the epistemic restriction. But letting this issue aside, and keeping the focus on considerations about usefulness, the epistemic restriction can still be argued for by following a the line of reasoning sketched next.

The epistemic restrictions imposed on scientific goals (and on the corresponding usefulness) by k(nowledge)-internal pragmatists can be supported on the basis of the hierarchic character of usefulness.[10] It seems clear that the attainment of certain goals depends on the attainment of some other goals. Since goals are usually accomplished by following certain procedures or performing certain practices which have proved useful to that end, a dependence between useful means steams from the dependence between goals. There is, thus, an objective sense in which it can be asserted that certain practices or beliefs are more useful than others. In this sense, it can be claimed that the most useful practices or beliefs would be those on which most useful practices or beliefs depend. In other words, the more dependence a certain useful practice generates in other useful practices, the more useful it is relatively to the others. It can be argued, on this basis, that epistemically useful practices are more useful than non-epistemic ones. To support this claim, it is, of course,

---

[10] The argument for the hierarchic character of usefulness is similar to that for the hierarchic character of ends, which can be traced back to Aristotle. In the opening passage of his *Nicomachean Ethics*, as he explains how different actions have different ends, he makes the following remark: "(. . .) then in all the cases the end of the master science is more worthy of choice than the ends of the subordinate sciences, since these latter ends are pursued also for the sake of the former. And it makes no difference whether the ends of the actions are the activities themselves, or something else additional to them, as in the sciences just mentioned" (Book I, Ch. I, 1094a). Aristotle explains the hierarchy of ends in terms of a certain dependence of some ends on the others. Attaining particular ends would always contribute to attain more general ones. The value of the first would always derive (and therefore depend), to some extent, on the value of the second, hence the superiority of the second over the first. The hierarchy of useful beliefs and practices asserted here is explained in similar terms. As the value of particular ends depends on the value of some more general ends, so the usefulness of some beliefs and practices depend on the usefulness of other beliefs and practices from which the usefulness of the former, to some extent, derive. It must be noticed that the dependency pointed out here steams from a kind of generality different from the one Aristotle appeals to. Both of them are, nevertheless, compatible, since one refers to theories and the other to practices. In the former, one thing is more valuable than other things if it includes the others, in the latter, if it is included in (or presupposed by) the others.

necessary to show how the accomplishment of non-epistemic goals depends on the accomplishment on epistemic goals. That is, it is necessary to show that success in goals like survival, adaptation, technological development, wealth creation and social welfare, depend on success in goals like observation, prediction, explanation, control, innovation and acquisition of true beliefs. It seems intuitively right that such dependence actually holds, being technological development, for instance, clearly dependent on innovation and control skills (even if an intense feedback occurs in this case, introducing certain complexity in the hierarchical relation).

2.2 Some Examples of Pragmatic Continuities Taken from the History of Astronomy

The fact that Babylonian observational records and parameters survived several scientific revolutions in astronomy is not usually emphasized in philosophical discussions. Yet Babylonian observations and parameters supplied an important source of empirical data in Ptolemy's *Almagest*, which in turn provided the main source of empirical data for Copernicus' *De Revolutionibus*.

In the history of astronomy, from the Babylonian period to the Copernican revolution in the renaissance, shared parameters were used in the description and prediction of the positions and motions of the celestial bodies. The arithmetical models of the Babylonian astronomers underlie different astronomical accounts originated from the Greek geometrical models. The transition from Edoxus' and Aristotle's concentric spheres universe, to Ptolemy's epicycle models, and, later, to Copernicus' heliocentric theory, occurred within the range of possible theoretical developments opened by the Babylonian observations and parameters.

2.2.1 Babylonian Continuities

2.2.1.1 Mathematical Notational Techniques

The sexagesimal system introduced by the Babylonians constituted the first place-value mathematical notation in history. It superseded the sign-value notation, which was the only one available before. Our actual decimal system is an extension of the sexagesimal system.[11] The latter, moreover, is still used in today astronomical

---

[11] Cfr. Neugebauer (1983, 159). The same author asserts in a previous work: "I think that this relationship between the Greek form of Babylonian astronomical computation and the older Hindu decimal number systems explains the creation of a decimal number system with place value notation, which was transferred by the Arabs to Europe and finally became our number system" (Neugebauer 1983b, p. 27).

computations and measurements. This is obvious in the sexagesimal division of time and angles in current instruments and tables. A base-60 system has the advantage that its base has a large number of conveniently sized divisors making easy calculations with fractions. The sexagesimal system is especially useful in measuring time, angles and geographic coordinates, and therefore is particularly important in astronomy.

The general computational achievement accomplished by the Babylonian astronomers is described by the historian O.E. Neugebauer in the following way:

> "The most important feature of this late-Babylonian astronomy is its mathematical character founded on the idea of computing the very complicated observed phenomena by addition of single components, each of which can be treated independently. Here for the first time in history we meet the fundamental method for the investigation of physical problems by using purely mathematical idealizations, a method which determined the course of all future science" (Neugebauer (1983b, p. 27).[12]

2.2.1.2 Observations and Mathematical Parameters Derived from Observations

Among the observational records, the zodiacal division of the ecliptic goes back to the Babylonian astronomy too.[13] The distinction between different constellations of stars lying along the apparent yearly path of the Sun across the heavens remained valuable as a frame in which make other observations related to celestial bodies or phenomena. Furthermore, Babylonian observations made possible to calculate the unequal intervals between equinoxes and solstices, as well as the length of the year. This, again, provides the basic observations for developing a minimal temporal-space frame in astronomy (Huxley 1964, pp. 3–13, p. 6). On the other hand, the discovery of the precession of the equinoxes (i.e., the gradual change in the longitude of the stars) is also attributed to the Babylonians. An especially important kind of observation made by the Babylonians is related to the eclipses. Eclipse records in Babylon were kept (according to Ptolemy) from the time of Nabonassar (747 BC).[14] These observations were later employed to determine the moon's eccentricity, latitude and epicycle radius (Neugebauer 1975, pp. 314–315).

Although the Babylonian arithmetical models for the motion of the heavenly bodies did not survived until Ptolemy, who, following the Greek tradition, rather used geometrical models, certain parameters derived from the former models did survive. For example, the arithmetical patterns for relating the variable length of daylight to the position of the sun in the ecliptic, which is important in geography,

---

[12] In *Astronomy and History. Selected Essays*, Springer-Verlag, New York, 1983, pp. 23–31.

[13] Cfr. Neugebauer (1983a, p. 161). George Huxley notices that the name of the zodiacal constellation (600 BC) constitute the earliest evidence of Babylonian influence upon Greek astronomy (Huxley 1964), pp. 3–13, p. 4.

[14] G. Huxley points out that the Babylonians were well aware of lunar eclipse cycles (Huxley 1964, pp. 3–13, p. 5).

for the determination of what later will be called "geographical latitude".[15] Among the surviving parameters are those for relations between planetary phenomena (like celestial bodies' relative positions and motions), and those for the anomalous motion of the moon and the planets. The calculation of the parameters for the different components of the lunar motion deserves separate mention, since the moon played a crucial role as a marker for celestial phenomena. According to O. Neugebauer, who is one of the most prominent historians of ancient astronomy, "the really significant contribution of Babylonian astronomy to Greek astronomy, in particular to Hipparchus' astronomy, lies in the establishment of very accurate values for the characteristic parameters of lunar and planetary theory and in particular in the careful preparation of the components of the lunar motion – longitude, anomaly, latitude and nodal motion. The value of one of these parameters, the evaluation of the length of the mean synodic month as 29:31, 50, 8, 20 days is not only fundamental for Hipparchus' theory of the moon but still appears in the Toledan Tables of the eleventh century'.[16] The determination of this parameter involved the application of the Babylonian mathematical theory of the lunar phases (Neugebauer 1983a, p. 158).

2.2.1.3 Measurement Procedures in Chronology

The Babylonian mathematical theory of the lunar phases made possible the mathematical articulation of the lunar calendar (Neugebauer 1983a, p. 158). It was possible to determine the different periodic variations which cause the intricate pattern in the variation of the time between consecutive new crescents.[17] A well developed lunar calendar allowed, in turn, the introduction of smaller units that could be employed as fundamental units of time measurement.[18] They did that by employing "lunar days" of exactly 1/30 of a mean synodic month. The coordination of the lunar calendar with the seasons of the solar year was also accomplished by introducing a definite intercalation cycle, which intercalates seven additional months in 19 lunar years.

---

[15] Cfr. Neugebauer (1983a, p. 166). This Babylonian procedure was later expanded by the Greeks by using a linear variation of the extremal length of day light but otherwise unchanged pattern. According to the resultant scheme one distinguishes between "climates" of equal length of daylight, arranged in the simple pattern of half-hour increment of the longest day. "(. . .), as a concept, the sequence of the climates of linearly increasing length of the longest day remained unchanged and dominated geographical lore from antiquity through Islam and the western Middle Ages", *ibid.*, p. 162.

[16] Cfr. Neugebauer (1983a, p. 163). In emphasizing the accuracy of the Babylonian, arithmetical models accounting for the moon's elements, G. Huxley remarks that Babylonian arithmetical progressions permitted to accurately predict lunar phenomena to within a few minutes of time, "The Interaction of Greek and Babylonian Astronomy", *cit.*, p. 7.

[17] Cfr. *ibid.*, p. 160.

[18] Cfr. ibid., p. 161.

Finally, it is worthwhile to mention that, according to Herodotus, the Greeks learned of sundials (which were used both as time pieces and as instruments for the determination of solstices) with their pointers and of the 12 parts of the day from the Babylonians (Huxley 1964, p. 5).

2.2.2 Continuities from Ptolemy

There are at least two successful and long-lasting procedural contributions that Ptolemy made in astronomy. First, the observations and mathematical parameters derived from the observations by applying both plane and spherical trigonometry.[19] Second, the cinematic models for the seeming motion of the heavenly bodies as seen from the earth, which constitute a good approximation of the geocentric cinematic models still applied today in astronomical experiments.[20] Besides that, there is also a remarkable mathematical advancement accomplished by Ptolemy that deserves at least a passing reference. His extension of plane trigonometry to spherical trigonometry broadened the range of possible calculations related to celestial phenomena. The chord was the only trigonometrical function used by Ptolemy. He had no equivalent for the cosine formula, but made use of straight lines perpendicular to the axes of the spheres, transforming the problem into one of solving only right angles (Toomer 1984, p. 7). Beyond the importance of this contribution in pure mathematics, it made possible the full development of a quantitative astronomy, the mathematical handling of observed motions. I will briefly comment on each contribution next.

2.2.2.1 Observations and Mathematical Parameters Derived from the Observations

Ptolemy's records of observations of the equinoxes were employed by Copernicus to review the variation in the length of the tropical year, and so to determine the nonuniform variation of parameters relevant for his solar theory. He relies too on Ptolemy's observations to derive the parameters for Venus and Mercury respective models

---

[19] Cfr. ibid., p. 164. For a clear description of the trigonometric procedures employed by Ptolemy in the *Almagest* see Toomer (1984), pp. 7–9.

[20] The later relevance of these two points has been emphasized in Swerdlow and Neugebauer (1984), p. 41. They actually highlight three points of Ptolemy's astronomy that consider of great importance for later astronomy in general and for Copernicus in particular. The first concerns the explicit use of observations for deriving numerical parameters. The second relates to implicit used of observations for the purpose of describing apparent motions and deriving the appropriate models. The third, which I have skipped given its lower pragmatic value in comparison to the former ones, concerns "the physical representations of the models that are supposed to exist in the heavens and produce the apparent motion of the planets".

(Swerdlow and Neugebauer 1984, p. 80). Moreover, Copernicus is able to confirm the decrease of both the earth's eccentricity and a nonuniform motion of the earth's apsidal line also by relying on Ptolemy's results, in this case on parameters such as eccentricity and nonuniform motion (Swerdlow and Neugebauer 1984, p. 73).

2.2.2.2 Cinematic Models for the Seeming Motion of the Heavenly Bodies as Seen from the Earth

As for the Ptolemaic cinematic models, two different aspects may be distinguished: one related to the assumption that celestial bodies have circular motions, another concerning the assumption that celestial bodies turn in circles around the earth. With respect to the first, it must be noticed that the construction of planetary orbits from circles, upon and within circles was not superseded until Kepler's constructions using ellipses showed more in agreement with observations. The radical consequences of Kepler's achievement, in particular, the abandonment of the Aristotelian principles of uniform and circular motion for the celestial bodies, have often obscured the fact that the Ptolemaic models (in which eccentricities were precisely determined) constituted good approximations of Kepler's. This approximative relation actually explains the similarity of cinematic models through different planetary theories. With regard to the second aspect, namely, Ptolemy's geocentric cinematic models, their pragmatic merits need also be emphasized. In making such emphasis, optical relativity must be recognized as a crucial factor not only to understand the development of the Copernican system, but also to assess the usefulness of the Ptolemaic models for such development. They were necessary for the development of the Copernican system, for they provided a mathematical description of the apparent motions of the celestial bodies as seen from the earth.[21] Since any accurate description of celestial bodies' real motion seems to require an accurate description of their apparent motion, it seems plausible to consider that Ptolemy's accomplishment of the latter made possible Copernicus accomplishment of the former.[22] In short, Ptolemaic cinematic models did not successfully account for the real motion of the celestial bodies, but they did so for their apparent

---

[21] "At not point, however did he [Copernicus] question the soundness of Ptolemy's models for representing the apparent motions of the planets, and so at no time did he carry out the sort of analysis that Ptolemy had, and that Kepler did later, to determine what really constituted an appropriate model for the planets" (Swerdlow and Neugebauer 1984, p. 77). The same idea appears again later, being expressed even more emphatically: "(. . .) Copernicus' object was to find physically permissible heliocentric models that would reproduce, as far as possible, the apparent motions of Ptolemy's models", p. 79.

[22] In his classical work (Kuhn 1979, pp. 100–101), T. S. Kuhn, after a thorough analysis of the historical episode, concludes that no new data prompted the astronomical revolution. According to him, Copernicus inherited Aristotle's and Ptolemy's astronomy. The former depended very much on the latter's observations, there were little new in his mathematics, and no better predictions. Furthermore, Copernicus offered a new mathematical description of the motion of the planets, but no physical explanation of that motion.

motion, and this proved useful independently of the cinematic system endorsed. The different ways in which optical relativity could affect astronomical observations were already known from very early times.[23] However, both astronomers and natural philosophers faced the practical impossibility to determine whether it actually played a role, and, if so, how it did. Such practical impossibility to determine how optical relativity affected astronomical observations did, of course, slow down for centuries the development of accurate cinematic models. But, on the other hand, it did not prevent the development of detailed and precise cinematic models reflecting the apparent motions of the celestial bodies as seen from the earth. As already said, considering these models just as models for apparent phenomena, their usefulness through different systems is clear. This is shown by the fact that the Copernican model maintained epicycles moving along the deferent (which explained retrograde motion in the Ptolemaic model) as a device to explain the apparent retrograde motion, and so, to deny retrogradation as a real phenomenon. Copernicus' replacing the equant with epicyclets kept his system dependent on Ptolemy's geometrical procedures to compute the apparent motion of celestial bodies (Swerdlow and Neugebauer 1984, p. 78).

More in general let us remember that, from the point of view of modern science, on which no absolute frame of reference is accepted, a geocentric frame is acceptable when its usefulness justifies that. A geocentric frame is useful today for everyday activities and most laboratory experiments. Whereas a heliocentric frame is most useful in solar-system mechanics and space travel. Galactic and extragalactic astronomy is easier if the sun is treated as neither stationary nor the center of the universe, but rotating around the center of our galaxy.

It should not strike us now that in his classical work, *The Copernican Revolution*, T. S. Kuhn, after a thorough analysis of the historical episode, concludes that no new data prompted the astronomical revolution. According to him, Copernicus inherited Aristotle's and Ptolemy's astronomy. The former depended very much on the latter's observations, there were little new in his mathematics, and no better predictions. Furthermore, Copernicus offered a new mathematical description of the motion of the planets, but no physical explanation of that motion.

The above examples are meant to show how effective procedures and useful observations remain valuable, even if they are not interesting any more for the development of current theories. Further, they may at some point prove useful again. As a final example to illustrate this last point, I will refer to a case that I find particularly revealing of the value of observations independently of the particular theories for whose development they were initially employed. I am referring to the fact that Babylonian observations have recently been used to reckon the change in

---

[23] Aristotle begins chapter 8 of *On the Heavens* with the following remark: "Since both the stars and the heavens as a whole are observed to change position, the change must occur either with both the heavens and the stars being at rest, or with both moving, or with the one moving and the other at rest" (*DC, Book Ii, Ch. 13 (289b1)*).

the gravitational acceleration of the earth (Newton 1974, pp. 99, 113–115). Friction in the tides is gradually slowing down the Earth's rotation about its axis and that is also decreasing the Moon's angular velocity in its orbit about the Earth. These effects are customarily described by saying that the lengths of the day and the month are both gradually increasing. Tidal friction is the only phenomenon that tends to change the length of the month. While other phenomena tend to change the length of the day, among them, effects on the Earth magnetic field and slow changes in the average radius of the Earth because of change in the average temperature of the Earth's interior. The respective accelerations of the Earth and the Moon, that is, their respective angular velocities, have both negative values. Which means that the respective rates of change for the angular velocity of the Earth's rotation and the orbital velocity of the Moon are both decreasing. The physicist Robert. R. Newton explains, in his article "Two uses of ancient astronomy", how nowadays ancient and medieval records of solar eclipses are used in the estimation of the accelerations.[24] Therefore, scientists today attempt to calculate the acceleration parameters for the Earth and the Moon on the basis of observations that have been recorded since around the eighth century BC.

References

Aristotle (IV B.C.). 1995. *On the Heavens* (trans: Stuart Leggatt). Warminster: Aris & Phillips.

Haack, Susan (ed.). 2006. *Pragmatism, old & new. Selected writings*. New York: Prometheus Books.

Hacking, Ian. 1983. *Representing and intervening. Introductory topics in the philosophy of natural science*. Cambridge: Cambridge University Press.

Huxley, G. (1964) The interaction of Greek and Babylonian astronomy. In *New lecture series* no. 16. Belfast: The Queen's University of Belfast.

Kuhn, T.S. 1979. *The Copernican revolution. Planetary astronomy in the development of western thought (1957)*. Cambridge: Harvard University Press.

Neugebauer, O. 1975. *A history of ancient mathematical astronomy, part I*. New York: Springer.

Neugebauer, O. 1983a. The survival of Babylonian methods in the exact sciences of antiquity and middle ages (1963). In *Astronomy and history. Selected essays*, 157–164. New York: Springer.

Neugebauer, O. 1983b. Exact science in antiquity (1941). In *Astronomy and history. Selected essays*, 23–31. New York: Springer.

Newton, R.R. 1974. Two uses of ancient astronomy. In *The place of astronomy in the ancient world*, ed. F.R. Hodson, 99–116. London: Oxford University Press.

Rescher, N. 1977. *Methodological pragmatism. A systems-theoretic approach to the theory of knowledge*. New York: New York University Press.

---

[24] "Ancient and medieval astronomical data allow us to form 25 independent estimates of the important acceleration parameter D'', at various epochs from about -700 to $+1300$. These estimates, combined with modern data, show that D" has had surprisingly large values and that it has undergone large and sudden changes within the past 2,000 years. It even changed sign about the year 800. The uncertainty in the value of D'' at any epoch from –700 to $+1300$ is about $2''$/century" (Newton 1974, p. 115).

Schurz, Gerhard. 1998. Kinds of pragmatisms and pragmatic components of knowledge. In *The role of pragmatics in contemporary philosophy*, ed. P. Weingartner, G. Schurz, and G. Dorn, 39–57. Vienna: Hölder-Pichler-Tempsky.

Suppes, Patrick. 1998. Pragmatism in physics. In *The role of pragmatics in contemporary philosophy*, ed. P. Weingartner, G. Schurz, and G. Dorn, 236–53. Vienna: Hölder-Pichler-Tempsky.

Swerdlow, N.M., & Neugebauer, O. 1984. *Mathematical Astronomy in Copernicus' De Revolutionibus*, Part I. New York: Springer-Verlag.

Toomer, G.J. (trans.) 1984. *Ptolemy's almagest*. London: Springer.

Wiener, Philip. 1973-74. Pragmatism. In *The dictionary of the history of ideas: studies of selected pivotal ideas*, vol. 3, ed. P. Philip, 551–570. Wiener: Charles Scribner's Sons.

Chapter 3
The Principle of Eurhythmy: A Key to the Unity of Physics

J.R. Croca

Abstract The unification of the basic physical laws, from classical physics, quantum physics to gravitic physics seems now possible. The key for this unity is the principle of eurhythmy, literally from the Greek, the principle of the most adequate path. It will be shown that Heron principle of minimum path, Fermat's principle of minimum time and de Broglie's guiding principle are no more than mere particular cases of the principle of eurhythmy. Furthermore, it will be shown, with concrete examples, from classical physics, quantum physics to gravitic physics, that all these branches of physics can be unified and understood in a causal way as particular cases of this general principle.

Keywords de Broglie guiding principle • Extreme principles • Fermat's principle of minimum time • Fundamental physics • Principle of eurhythmy • Quantum physics • Unity of physics

3.1 Introduction

The aim of Physics has always been the Unity. This ideal means that physicists look for a very basic principle, from which it would, at last in principle, be possible to derive all particular laws for describing the physical reality at different scales of observation. Presently in physics we are faced with two independent, even opposite, domains, the said classical domain and the quantum realm. The principle

---

J.R. Croca (✉)
Departamento de Física, Faculdade de Ciências, Centro de Filosofia das Ciências,
Universidade de Lisboa, Campo Grande Ed. C8, 1749-016 Lisbon, Portugal
e-mail: croca@fc.ul.pt

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology, and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_3,

of eurhythmy[1] comes from the Greek ***euritmia***, which is the composition of the root *eu* plus *rhythmy*. With *eu* standing for the right, the good, the adequate, and *rhythmy*, for the way, the path, the harmonic motion. The composed word meaning: the adequate path, the good path, the good way, the right way, the golden path, and so on.

In the present work I intend to show that the principle of eurhythmy is, indeed, a basic key for the understanding of the whole physical world.

3.2 Some Basic Principles

One of the first principles for understanding the physical world was discovered by Heron of Alexandria back in the first century A.D. In order to explain the reflection of light (Fig. 3.1) and derive the respective law Heron established the principle of the minimum path. This principles states that light coming from a point like source S and going to the observer O, after being reflected by the mirror, from the multiple possible paths follows the shorter one (Fig. 3.2).

Fig. 3.1 Reflection phenomena

---

[1] The Greek name eurhythmy for the basic principle of Nature was suggested to me by my dear friend Professor Gildo Magalhães.

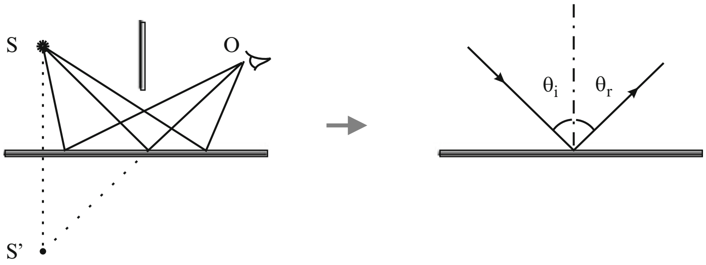
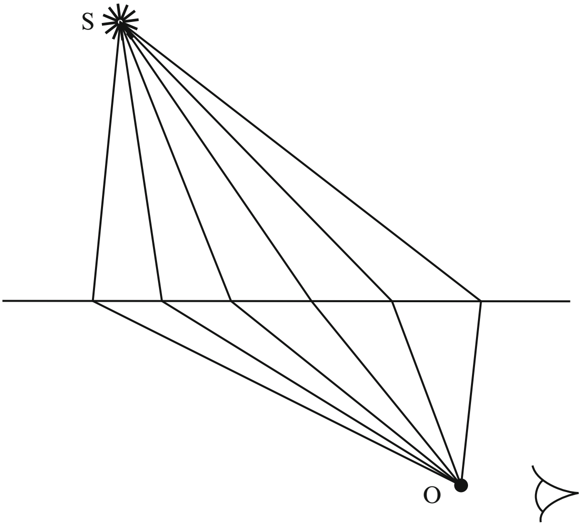


Fig. 3.2 Heron principle of shortest path

Fig. 3.3 The principle of minimum time



From this principle Heron derived (Eckt and Zajac 1974), by geometrical considerations, the reflection law stating that the angle of reflection equals the incident angle, $\theta_i = \theta_r$.

About a millennium and half later, Pierre de Fermat discovered the principle of minimum time. With the help of this principle Fermat was able to derive Snell's empirical law for the refraction of light. This law describes how light propagates in different optical media. The principle of minimum time states that from all possible paths, from S to O, see Fig. 3.3, light follows the path such that the time taken in the whole course is minimum.

This principle of minimum time says that the light coming from a point like source, the point S, in order to reach the observer at point O, after travelling through two different optical media, follows a path such that the time taken in the whole course is a minimum.

That is, light behaves in such a way that it "chooses" from among all possible paths the one that takes less time. From this principle Fermat derived Snell's law for the transmission of the light in different optical media. For the detailed calculation see Appendix A1. The basic idea behind this principle of minimum time is that the light that goes from point S to reach point O, after travelling through different optical media, where it has different velocities follows, from all possible paths, the one that takes less time. In short, this natural entity we call light “chooses” the best, the most adequate path, that is, it follows the principle of eurhythmy.

In these conditions light seems to behave much like the human beings when they have to go from on point to another. Suppose you happen to be in the centre of Lisbon, at the Chiado for instance, and want to go to Almada city, at the other side of the Tagus River. The shortest way to go is by the 25 de Abril bridge. Nevertheless, we are in the summer time and too many people want to go to the beaches to enjoy the sea and the Sun so, at the moment, there is one of the biggest jams at the 25 de Abril bridge. Aware of these facts, instead of trying to go through the shortest path, by the 25 de Abril bridge, you went by Vasco da Gama bridge. You made this choice because even if this alternative path is three or four times longer you arrive sooner at you destination, the city of Almada, because you can travel much faster.

In the case of the light it travels the shortest possible path in the medium where its velocity is slower, and the longest path in the medium in which it can travel faster so that, at the end, it takes the less possible time on going from one point to the other.

Naturally, Heron principle of the shortest path is a particular case of Fermat’s principle of minimum time. This is easily seen because in the reflection phenomena the light travels always in the same optical medium and with the same velocity. In such conditions, the shortest path is the one that takes less time.

The principle of minimum time of Fermat was soon followed by Maupertuis’s principle of least action for classical mechanics. These principles were and still are commonly named extreme principles. It was shown (Rica da Silva and Croca) that classical mechanics can be derived from these principles of extreme. Since these principles are mere particular formulations of the same basic principle, it is only natural to assume that classical physics is nothing more than a consequence of principle of eurhythmy.

3.3 Nonlinear Quantum Physics

At the beginning of the twentieth century physicists were faced with two apparently contradictory ways for the observed behaviour of nature (Croca 2003). At the quantum scale of observation things seemed to occur in a different and even strange way. A quantum entity, such as an electron or so, shows on the one hand corpuscular properties of localization and by the other hand extended wave-like properties.

In order to make things a little more concrete and size the real problem let us look briefly at the well-known double slit experiment.

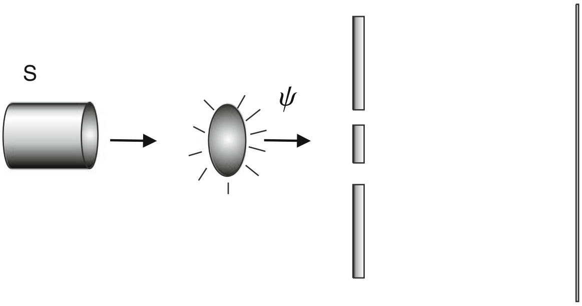


Fig. 3.4 Double slit experiment

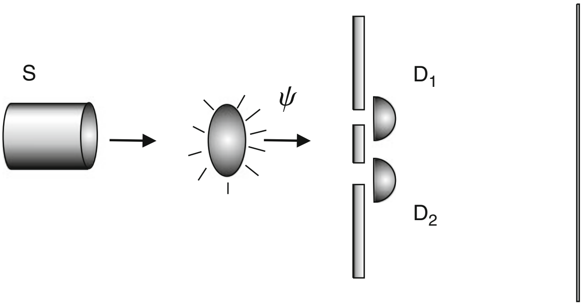


Fig. 3.5 Double slit experiment with two detectors

3.3.1 Double Slit Experiment

Since the beginning of quantum physics the double slit experiment has played a most important role in the conceptual development of quantum physics. Consequently this experiment has been done and redone many times, with different quantum entities, like electrons, photons, neutrons, atoms, molecules, and so, giving always the same result.

The experiment runs like this: A source of quantum entities, of electrons for instance, emits one at a time. This means that in the whole apparatus only one and only one single electron is present at each instant of time. This basic requirement is made in order to avoid any possible interaction with other electrons.

Let us consider this fundamental experiment, sketched in Fig. 3.4.

The electron in its way encounters a screen with two slits. If one places a detector right behind each slit what is to be expected (Fig. 3.5)?

It happens that sometimes one detector is triggered and at other times it is the other detector that is activated. In any case, and here is the point, it never happens that the two detectors are activated at the same time. Naturally, this situation could never be expected to occur since we are dealing with one single electron at a time. In the whole apparatus, for each event, there is only one single quantum particle. If this basic conceptual and practical requirement is not met the experiment is meaningless. Therefore, from this first analysis the observer would conclude that the electron would go through one slit or the other.

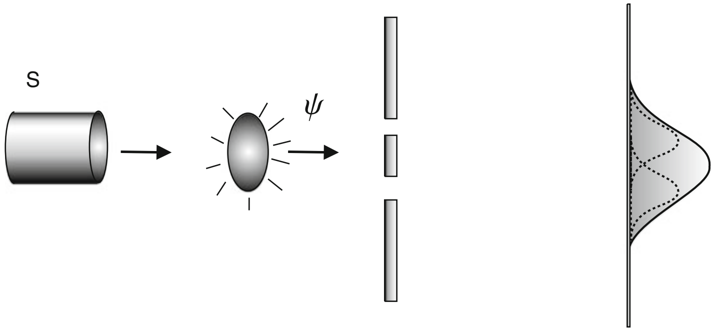


Fig. 3.6 Double slit experiment with corpuscles

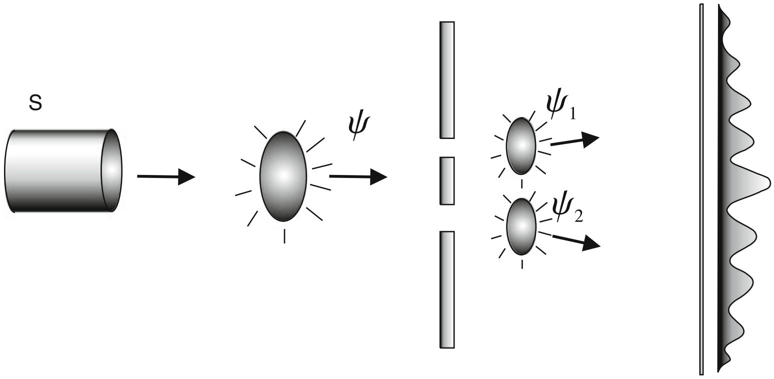


Fig. 3.7 Double slit experiment

Now, if one removes the two detectors from the slits, and places a large array of detectors, sufficiently far away in the route of the particles that went through the slits, what is to be expected? What shall be the distribution of impacts of the electrons on the array of detectors?

If one accepts the previous conclusion that the electron went by one or the other of the slits, then one would expect a continuous bell shape gaussian distribution of impacts on the detector, see Fig. 3.6.

That is, if the electron behaves only like a corpuscle entity the observed distribution would result from the contribution of electrons that sometimes come from one slit, plus the contribution of the corpuscles that, at other times, come from the other slit.

Still, experimental evidence clearly shows that instead of a continuous gaussian pattern, what in fact is observed is an interferometric pattern distribution as shown in Fig. 3.7.

In order to explain the observed interference pattern one has to assume that the quantum particle, in this case the electron, has somehow gone both ways. An interference pattern is always a result of the superposition of at least two waves. This means that the quantum entity we call electron went through both slits. Under these conditions we arrive at the conclusion that the electron must be a wave.

Now we have a problem!

The first experiment, with the detectors placed behind the two slits, indicates that the electron is a kind of localized entity, that is, a corpuscle that went though one slit or through the other.

The second experiment leads us to conclude that the electron is an extended entity, a wave, because it went through both slits at the same time giving origin to an interference pattern.

Therefore the quantum entity we call electron must have a very weird behavior: It has to pass through one slit **or** the other and also must go, at the same time, through **both** slits.

Summarizing: In the single particle double slit experiment the quantum particle must pass:

1. through one slit **or** the other slit.
2. through one slit **and** the other slit

How can this be possible! Is it possible to find an explanation for this apparent logical contradiction for the behavior of nature at the quantum level?

3.3.1.1 Orthodox Interpretation

When we have a very tough problem to solve, in general, the first and even the easiest approach evolves a process which, in the whole, is very similar to those used by mankind in all times, since the beginning of history till now. When faced with the very complex problems posed by the everyday life, due to the enormous difficulty in finding a correct natural, causal and rational explanation, the first approach is, in general, to get help from what is usually named, the unknown, the occult, the transcendental, the mysterious and the weird forces of nature. First one looks for some kind of regularities, pattern repetitions, then tries to formulate some first rough empirical laws. The next step is finding some kind of justification for them. Most times, as we well know, mainly due to the difficulty in finding a causal rational explanation, or even by an inner desire, these first pseudo-justifications involve Godlike transcendental justifications. If it was found that at a certain place and at a certain epoch of the year it usually rains, then it certainly was because that was the will of the Gods or of other similar powerful transcendent entities. Trying to find a causal and rational explanation for the natural phenomena was, in the best case, a pure loss of time because there was nothing more to look for. To solve the problem of the dualism wave-corpuscle posed by nature, at the quantum scale of observation, the first idealistic non-causal approach for finding a solution was made much in a similar way.

In order to conciliate the two apparently contradictory affirmations resulting from the complex quantum behavior, Niels Bohr developed in 1927 his famous principle of complementarity. This principle assumes that a quantum particle has two dual and opposite natures. Sometimes quantum entities show the undulatory extended nature, other times the corpuscular localized properties. The two properties never showing at the same time. According to the experiment some times we observe the corpuscular localized properties other times undulatory extended properties. These two opposite properties are a direct consequence of Fourier ontology (Croca 2007). This ontology claims that only the infinite in time and space harmonic plane waves do have a well definite frequency. All other finite waves are necessarily a composition of infinite harmonic plane waves. In this conditions the more a quantum entity is localized de more harmonic plane waves are necessary to build it. In the limit when the position of a quantum entity is known with absolute precision, the number of harmonic plane waves necessary to make that structure is infinite. This means that if we know the position of a quantum particle without error them, the same very single particle has a infinite number of frequencies and consequently possesses all possible energies. Inversely, if the energy of the particle is known without error, we are in the other extreme. In such conditions, only one single harmonic plane wave is necessary to describe it. In such conditions the position of the quantum particle is completely unknown because the quantum entity, occupies indeed all space and time.

So, in this idealistic approach, the double slit experiment undoubtedly shows those two opposite properties of quantum systems that manifest clearly one at a time:

(a) When the detectors are placed in front of the slits and only one of the two is fired, the quantum particle exhibits its corpuscular nature.
(b) With the detectors removed, the electron passes through both the two slits showing the undulatory feature in order to produce the observable interference pattern.

Therefore, in the Copenhagen, or Bohrean, interpretation of the quantum physics, the double slit experiment can be interpreted in a general idealistic conceptual framework in the following way:

The quantum particle is to be described by a probability wave, containing all information on the quantum system. This wave goes in direction of a screen with two slits. There the initial wave splits into two probability waves one coming from slit one the other from slit two. These two waves, coming from the slits, are directed to the detection region where they overlap. The detection region is usually composed of a large array of detectors. When the composite wave, which is the sum of the two, arrives at the detection zone one small detector from the array is triggered. The distribution of the clicks, corresponding to the successive arrivals of electrons, follows an interference pattern distribution given by the squared modulus of the total probability wave.

Therefore in order to explain and predict the result from the double slit experiment Bohr denies the very existence of the objective reality. Before measurement, the entity we call electron, does not exists really. What exists is only a bunch of two potential electrons, one coming from slit one the other from slit two. One of these potential electrons can eventually be made real by the act of measurement. The whole conceptual orthodox theoretical construction is made upon these idealistic ideas.

3.3.1.2 Causal Interpretation

Even if it seems difficult to understand the apparent strange behaviour of the quantum entities a causal, rational, beautiful and even intuitive, explanation for the apparent logical contradiction posed by the double slit experiment is possible.

Contrary to the Bohrean proposal this explication starts from the basic assumption that there is an objective reality from which the observer is a part. In order to coherently follow this causal program it is necessary, from the very beginning, to deny Fourier ontology. In these circumstances we accept as natural that a finite wave, after all real waves are finite, may have a well definite frequency. So a single quantum particle which position is known with good precision can, consequently, have a single value for the energy. The mathematical tool that allows us to build this causal model only recently was developed and is called local analysis by wavelets (Hubbard 1998; Chui 1992).

It was the great physicist Louis de Broglie, in the twenties of the twentieth century, the one who advanced the first rational explication for the riddle posed by the double slit experiment. De Broglie started from the assumption that a full explanation for quantum phenomena needed a nonlinear theory. Still this nonlinear theory could be approached, in certain cases, at the formal level to the orthodox linear theory.

De Broglie (de Broglie 1964; de Broglie and Andrade e Silva 1971) proposed that the thing we call a quantum particle, implicitly assumed to be a single point like structure, is indeed a very complex entity. In his view a real quantum particle is composed of an extended, yet finite, part, the wave, plus a very localized and indivisible structure, the corpuscle. Later (Croca 2003) this early model was developed and the quantum particle is now mathematically described by a wavelet solution to a nonlinear master equation that describes both extend and localized properties.

Commonly, this wave is named theta wave, or even pilot wave, and the localized structure singularity or the corpuscle. The following drawing, Fig. 3.8, tries to picture a causal quantum particle.

For all practical purposes the energy of the quantum particle is the one of the singularity. The energy of the theta wave is so small that the common detectors are unable to see it. Now a question arises! How are the wave and the singularity related?

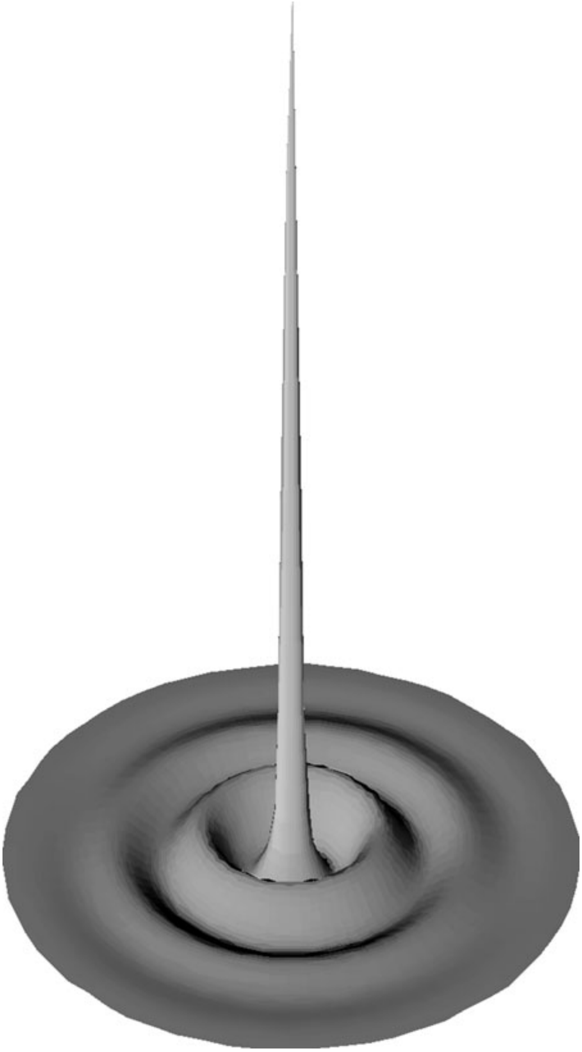


Fig. 3.8 Graphic representation of a quantum particle

3.4 The Principle of Eurhythmy

The answer to the last question is given by the principle of eurhythmy. Initially, the relation between the theta wave (pilot wave) and the singularity was named by de Broglie as the guiding principle. Meaning that the singularity is guided, through a nonlinear process, preferentially to the regions where the theta wave has higher intensity. Still this name does not give full account of the complex physical process involved. What really happens is that the singularity, or the corpuscle, necessarily immersed in the theta wave field when moving, randomly from one point to another, follows, in average, the best possible path. The best possible path is the one where

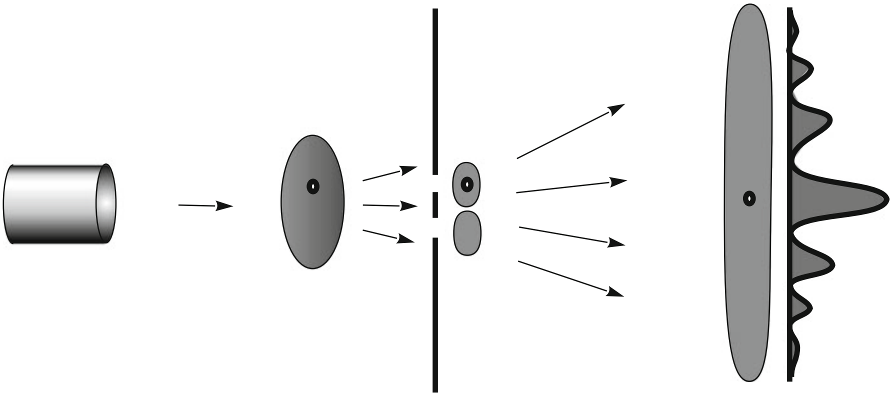


Fig. 3.9 Double slit experiment explained by the principle of eurhythmy

the theta wave has higher intensity. So, in reality it is not a simple action of guiding. The theta wave naturally guides the corpuscle, but this guiding action is of a very special kind. The theta wave, in fact, guides the singularity to the best possible path, the average path where the intensity of the theta wave is greater. The singularity avoids the regions where the intensity of the theta field is null because in these regions its very existence is in danger, in the sense that in such zones the corpuscle needs to regenerate the surrounding field at the expenses of its own energy. Therefore the motion of the corpuscle in the theta wave field follows always the best possible path, the most adequate path, that is, the average motion follows the principle of eurhythmy. It is assumed, of course, that any corpuscle, as long as it exists, has an inner energy of its own and furthermore keeps always in motion with its natural velocity. The corpuscle moves incessantly, in the theta wave field, with an instantaneous huge velocity, called the natural velocity. Nevertheless, due to the chaotic nature of the subquantum field the average velocity, which is the observed velocity, can go from zero to the natural velocity.

In these circumstances the principle of eurhythmy unifies by the one hand the principle of minimum time of Fermat and the guiding principle of de Broglie, by the other hand it gives them their true natural deep meaning.

Let us now see briefly how the principle of eurhythmy explains in a clear and beautiful way the double slit experiment.

Figure 3.9 sketches the experiment where, as always, the source emits a single quantum particle composed, as we have seen, of a wave plus the corpuscle.

When the particle arrives at the screen with the slits the extended theta wave field crosses the two slits simultaneously. The very much localized and indivisible corpuscle follows through one slit or the other going immersed in one theta wave or in the other. These two theta waves, one with the corpuscle, the other without it, in theirs natural courses spread in such a way that they overlap, giving origin to a total single theta wave in which the corpuscle is immersed. This total theta wave, being the composition of the two individual theta waves coming from the slits has, as we well know, a distribution of intensity with an interferential form.

Now, the principle of eurhythmy tells us that the corpuscle, among all possibilities, “chooses” to be localized preferentially at the regions of the space where the intensity of the theta field, the total theta wave, has higher intensity. In these conditions, the first corpuscle arrives at the detection region giving origin to a single click. The position of the click follows naturally the principle of eurhythmy, that is, is localized preferentially in the regions where the intensity of the total theta wave is greater. The next arriving corpuscle follows precisely the same principle. The next do the same. So, as the time goes, the distribution of clicks, in the target detector, starts gradually getting a net form. In such conditions, at the end, a clear and stable interference pattern becomes perfectly visible. This interferometric pattern is, of course, a perfect copy of the form of the total theta wave intensity field.

In this way, the apparent logical contradiction faced by physicists at the first quarter of the twentieth century for the quantum particle needing to pass through:

1. One slit **or** the other
2. One slit **and** the other

is satisfactorily solved in a most unambiguous way.

This natural explanation, beautiful and intuitive, can be summarized in the following manner:

1. The indivisible corpuscle passes through – One slit **or** the other.
2. The extended theta wave passes through – One slit **and** the other.

As we have just seen it is perfectly possible to comprehend the apparent mystery posed by nature at the quantum level of description without any need to get help from transcendental or supernatural forces and furthermore without any necessity of denying the existence of the objective reality. What was necessary, to solve the problem, was imagination and above all the believe that mankind can, progressively, unfold the apparent mysteries posed by nature.

Naturally the predictions for results in the double slit experiment can be reached either in terms of the linear orthodox theory or by the more general nonlinear causal theory. Only in certain careful designed experiments, de predictions of the two theories are different. Still, in the working of some super resolution microscopes, the orthodox linear theory has already shown its limits of validity. The working of these new super microscopes need to be understood the nonlinear quantum theory, namely the generalized uncertainty relations.

The previous treatment of the principle of eurhythmy was mostly of descriptive nature, to see the detailed mathematical derivation see Appendix A2.

3.5 The Principle of Eurhythmy and Gravitation

Even before the time of Newton (Edwards 2002; Evans 2002) many efforts were made with the aim of explaining the gravitic interaction. Namely, the early efforts made by Kepler, and later by Descartes who assumed that planets were carried

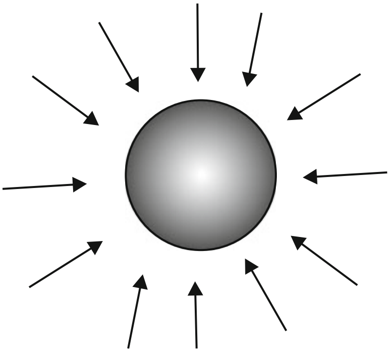


Fig. 3.10 An isolated body does not move due to the conjugated action of the pushing force

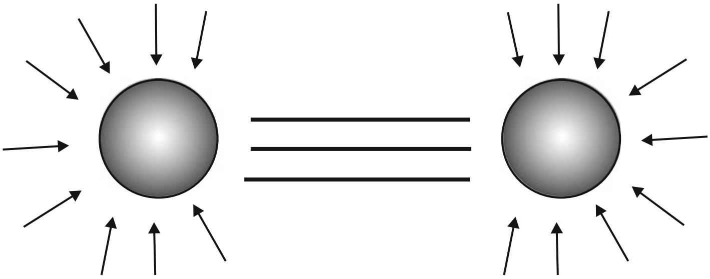


Fig. 3.11 Two bodies approach due to the mutual shielding action

by vortex (tourbillons) of a celestial fluid. Huygens was the first to propose a mechanism for the gravitation that was supported by calculation. Nicholas Fatio, a friend of Newton, proposed, in 1690, to the Royal Society of London a corpuscular theory of gravity. Nevertheless it was Georges-Louis Le Sage (1724–1803) who proposed the most developed theory to explain gravity.

We must point out that the laws for describing **how** gravity forces act were known mainly due to the work of Newton. Now the problem to solve was the **why** of these forces. That is, what was behind the known observed behaviour of such natural forces?

Le Sage tried to explain the forces of gravity in terms of his ultramundane corpuscles filling all space and striking the gravitic bodies impinging on it a pushing force. An isolated body would not move, see Fig. 3.10, due to the symmetric conjugated action of equal and opposite pushing forces of the ultramundane corpuscles. Nevertheless two bodies, Fig. 3.11 each one making a kind of shield to each other so that at the end the net effect was that they approach each other. It is possible to show that this overall pushing force acts according to Newton attraction law of the inverse of the square of the distance. This theory of Le Sage knew a certain amount of success till the late nineteenth century. Then, it was shown, using statistical physics, that this theory had a big problem with the principle of conservation of energy. In the second half of the twentieth century there were made various attempts of revival of Le Sage pushing theory using either

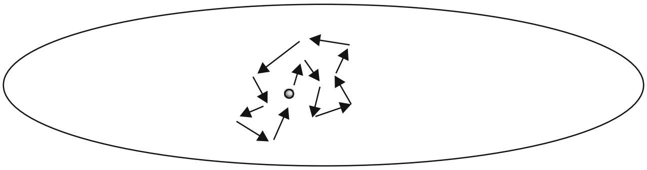


Fig. 3.12 Gravitic particle. The singularity moves chaotically in the theta wave field preferentially to the points of higher intensity following the principle of eurhythmy

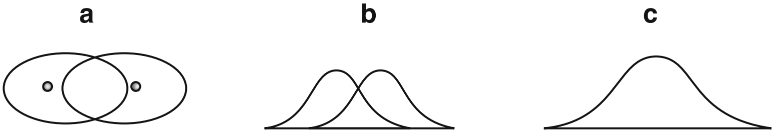


Fig. 3.13 Overlapping of two single gravitic particles, (**a**). Section of the radial average intensity of the individual theta waves, (**b**). Overall average intensity of the theta field resulting from the composition of the two theta fields, (**c**)

corpuscles or waves. Still, all those attempts had problems with the principle of conservation of energy, even if some authors tried to devise ingenious ways of overcoming the problem. In reality none of those gravitational theories based on a crude linear force impinging on the gravitic bodies worked well. To make the thing worse all those theories lacked a unitary approach for explaining physical reality.

Now thanks to the principle of eurhythmy the route for understanding the gravitational phenomena is open.

The basic idea is very simple. Just like in nonlinear quantum physics the concept of quantum particle means a multifaceted body, in gravitic physics the fundamental gravitic particle is also a very complex entity. The concept of particle in this new unitary physics involves implicitly a chaotic interaction between the singularity and the theta wave described by the principle of eurhythmy. The gravitic basic particle also shares the same fundamental nature. Thus the fundamental gravitic particle is composed of a wave plus a singularity, the graviton. The gravitic corpuscle being immersed in the field of the theta wave is subject to a permanent chaotic motion. This motion proceeds according to the principle of eurhythmy. That is, the graviton chooses to be preferentially at the points where the theta gravitic field has higher intensity. In this conditions an isolated single graviton immersed in its gravitic theta wave, due to the radial symmetry of the field in all directions, has also equal probability of moving in all directions, therefore its mean velocity is equal to zero. Even if the graviton moves with a huge instantaneous velocity in each chaotic jump, its mean velocity is, nonetheless, zero, see Fig. 3.12.

When we have two gravitic particles, occupying partly the same region of space, their respective theta waves overlap, see Fig. 3.13.

The average theta wave field is the result of the composition of the two single theta fields. In such circumstances, the two gravitons, immersed in the total theta field, each one moves preferentially to the points where the field has higher intensity, according to the principle of eurhythmy. Consequently due to the distribution of the total theta field, which is a composition of the two individual theta waves, the two gravitons, approach each other. The form of the total theta field, as can be seen from the drawing Fig. 3.13c, has intensity greater along the line connecting the two single theta waves. Using the old classical language we would say the two gravitons attract each other. That is, due to the nature of the resulting theta wave field, in average the gravitons move in a direction that leads them to approach each other.

The common gravitic bodies are no more than a huge composition of many gravitic particles. The composition of this large number of individual theta waves gives rise to a total theta wave field where the gravitons, the gravitic singularities, move according to the principle of eurhythmy. In fact, assuming that the total average theta field has approximately a linear variation in the neighbourhood of the Earth, it is possible to derive the law of Galilee. This law says that a body falls to the Earth in such a way that the travelled spaces are proportional to the squares of the time taken. In order to see the full calculation look at Appendix A3.

3.6 Conclusion

It was shown that the principle of eurhythmy is indeed the very basic conceptual tool for the understanding of the whole physical world in a natural causal way. In reality this basic principle promotes the very unity of all branches of physics. All physical theories, classical physics, including electromagnetism, quantum physics, and gravitic physics, are only particular cases of the application of the principle of eurhythmy. Still, we ought to be aware that the principle of eurhythmy does not in any way goes against the principle of conservation of energy, on the contrary. Yet this principle is a step ahead in the comprehension of the natural phenomena. There is a much more fundamental and important level of order in Nature described precisely by the principle of eurhythmy.

In this natural causal physics, valid at the diverse scales of observation, united by the principle of eurhythmy, there is a kind of average determinism. Nevertheless and in reality, this does not stand for a complete determinism. At the level of single individual, or even few, physical entities there are always an omnipresent unpredictable chaotic motion. Therefore, as a natural consequence, at the level of the single or few entities there is no determinism. We cannot predict the behaviour of the single entities. What we can predict is simply the overall behaviour of a large multiplicity of physical entities.

Appendix

A1. Derivation of Snell's Refraction Law from Fermat's Principle of Minimum Time

This derivation can be found in almost all good textbooks on optics. Nevertheless for the sake of reference we shall present here a simplified version. In order to do that, consider the following sketch where the light is emitted from point S and later is detected at point P.

The time t taken by the light in the course from point S to point P is equal to the time in the incidence region t_i plus the time in the transmission medium t_t,

$$t = t_i + t_t. \tag{A1.1}$$

These times being given by the travelled path divided by the velocity, which must be expressed in terms of the variable x,

$$\begin{cases} t_i = \dfrac{\bar{S}\bar{O}}{v_i} = \dfrac{1}{v_i}\left(x^2 + y_i^2\right)^{\frac{1}{2}} \\ t_t = \dfrac{\bar{O}\bar{P}}{v_t} = \dfrac{1}{v_t}((h-x)^2 + y_t^2)^{\frac{1}{2}} \end{cases} \tag{A1.2}$$

Since the total possible times is now a function of the variable x, $t = t(x)$, in order to find the minimum of all those possible times we use the habitual calculus technique

$$\frac{dt(x)}{dx} = 0, \tag{A1.3}$$

which gives the condition

$$\frac{1}{v_i}\frac{x}{(x^2 + y_i^2)^{\frac{1}{2}}} = \frac{1}{v_t}\frac{(h-x)}{((h-x)^2 + y_t^2)^{\frac{1}{2}}}$$

or

$$\frac{1}{v_i}\frac{x}{\bar{S}\bar{O}} = \frac{1}{v_t}\frac{(h-x)}{\bar{O}\bar{P}}. \tag{A1.4}$$

From the sketch Fig. A1.1 we see that

$$\sin\theta_i = \frac{x}{\bar{S}\bar{O}} \quad \text{and} \quad \sin\theta_t = \frac{(h-x)}{\bar{O}\bar{P}}, \tag{A1.5}$$

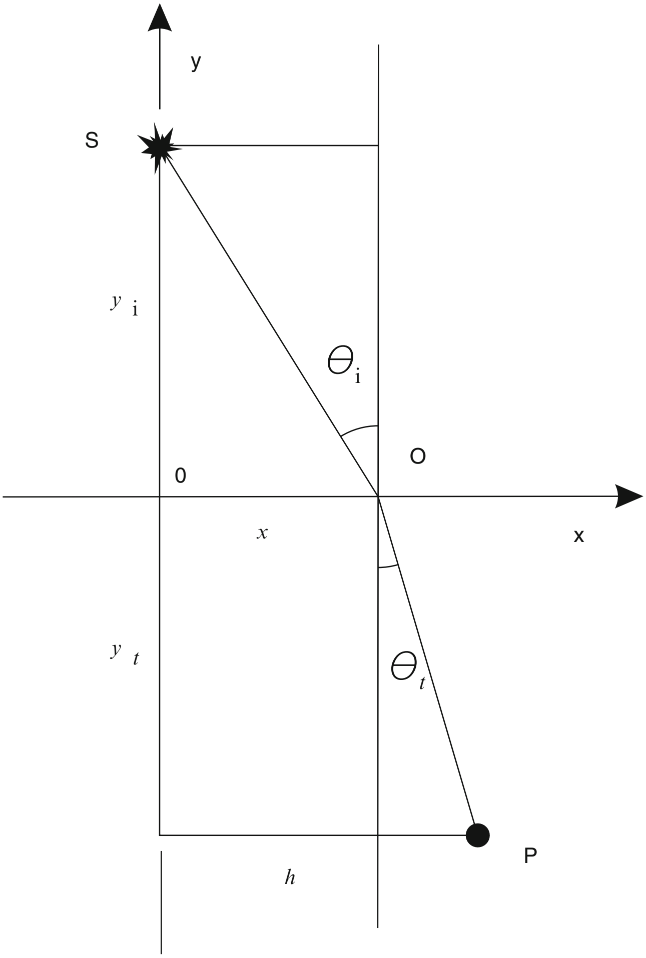


Fig. A1.1 Schetch for derivation of Snell's law from Fermat principle of minimum time

which by substitution in (A1.4) gives

$$\frac{1}{v_i}\sin\theta_i = \frac{1}{v_t}\sin\theta_t, \tag{A1.6}$$

or, multiplying by the constant c, it gives finally Snell's refraction law

$$n_i \sin\theta_i = n_t \sin\theta_t, \tag{A1.6$'$}$$

where $n_i = c/v_i$ and $\mathrm{n}_t = c/v_t$ are the indexes of refraction in the incident and transmission mediums.

A2. Derivation of Snell's Refraction Law from the Principle of Eurhythmy

In the derivation of Snell's refraction law from the principle of Eurhythmy we start from the following assumptions:

Basic:

1. The space available to the corpuscle is only and only the theta wave field. No corpuscles are found where the theta wave field is null.
2. In this approach there is energy conservation of the corpuscle. More complex cases where energy dissipation is important are not considered here.
3. The natural velocity of the corpuscle v_N is enormous, $v_N >>> c$. Still the observed velocity, the average, velocity can be very small even null.
4. The corpuscle moves preferentially to the zones where the theta wave field has higher intensity.

Formal:

1. The space available to the corpuscle is divided into cubic cells of linear size ℓ_0.
2. In each transition, which takes a very short time t_0, the corpuscle may move only to the adjacent cells.
3. The explicit form of the probabilities of transition are, for all practical purposes, the mathematical expression of the principle of eurhythmy.

In this context the natural velocity has the form

$$v_N = v_0 = \frac{\ell_0}{t_0} >>> c. \tag{A2.1}$$

where c is the habitual constant standing for velocity of the light in a medium with index of refraction one.

It is convenient to make a sketchy representation of the space in order to write the explicit form for the probabilities of transition. In two dimensions we have Fig. A2.1 where $p_{i\to i+1,j}$ corresponds to the probability of transition from cell $C_{i,j}$ to cell $C_{i+1,j}$ and $p_{i\to i,j}$ is the probability of remaining in the same cell $C_{i,j}$. Since there is no danger of confusion for notation simplicity reasons the index not corresponding to the transition is dropped. So in the following we make $p_{i\to i+1,j} \equiv p_{i,i+1}$, $p_{i\to i,j} \equiv p_{i,i}$, and successively. For the other probabilities their meaning can be inferred from the sketch. Naturally, for the generic cell $C_{i,j}$ the sum of the exit probabilities plus the probability of remaining must be one, that is

$$p_{i,i-1} + p_{i,i+1} + p_{i,i} + p_{j,j-1} + p_{j,j+1} = 1. \tag{A2.2}$$

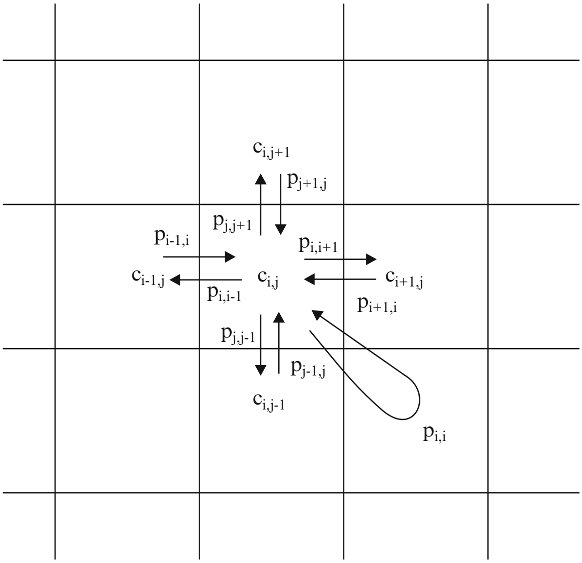


Fig. A2.1 Generic cell in a two-dimensional field with transition probabilities

Note that for notation simplification the index *j* was dropped in the horizontal representation and *i* in the vertical. As the following To explicit the analytic form of the probabilities of transition it is convenient to consider the following cases:

1. The corpuscle is in an initial condition such that is without any preferential direction for the motion. In this case the form of the probabilities stems directly from the principle of eurhythmy giving

$$p_{i,i-1} = \frac{I_{i-1,j}}{I_{i-1,j} + I_{i+1,j} + I_{i,j} + I_{i,j-1} + I_{i,j+1}} \tag{A2.3-1}$$

$$p_{i,i} = \frac{I_{i,j}}{I_{i-1,j} + I_{i+1,j} + I_{i,j} + I_{i,j-1} + I_{i,j+1}} \tag{A2.3-2}$$

$$p_{i,i+1} = \frac{I_{i+1,j}}{I_{i-1,j} + I_{i+1,j} + I_{i,j} + I_{i,j-1} + I_{i,j+1}} \tag{A2.3-3}$$

$$p_{j,j-1} = \frac{I_{i,j-1}}{I_{i-1,j} + I_{i+1,j} + I_{i,j} + I_{i,j-1} + I_{i,j+1}} \tag{A2.3-4}$$

$$p_{j,j+1} = \frac{I_{i,j+1}}{I_{i-1,j} + I_{i+1,j} + I_{i,j} + I_{i,j-1} + I_{i,j+1}} \tag{A2.3-5}$$

   where $I_{i,j}$ is the intensity of the theta wave in the cell $C_{i,j}$ and respectively.

2. The corpuscle due to some initial preliminary process in the surrounding theta wave field behaves as if it has a preferred tendency for motion.

In this case the expression for the probabilities of transition need to have two terms:

(a) One due only to the principle of eurhythmy depending only on the relative intensity of the field.
(b) The other resulting from the direct modification of the field in the neighborhood of the corpuscle, which usually is traduced by the so called initial velocity. In this approach the absolute value of the initial velocity is assume constant.

In such conditions for the generic form of the probabilities we got

$$p = \tilde{p} + p_0, \qquad p_0 = \text{constant} \tag{A2.4}$$

where $\tilde{p}$ is the direct contribution of the principle of eurhythmy and p_0 resulting from the initial velocity. In this situation the equation of conservation of probability (A2.2) assumes the form

$$\tilde{p}_{i,i-1} + q_{oh} + \tilde{p}_{i,i+1} + p_{oh} + \tilde{p}_{i,i} + \tilde{p}_{j,j-1} + q_{ov} + \tilde{p}_{j,j+1} + p_{ov} = 1, \tag{A2.5}$$

with q_{oh} corresponding to the term for the initial motion backwards, p_{oh} to the motion forward, for the vertical initial component, q_{ov} downwards and p_{ov} upwards. Naturally

$$q_{oh} + p_{oh} + q_{ov} + p_{ov} = \text{const} \tag{A2.6}$$

with each one also constant.

The master stochastic equation describing the number of the corpuscles in the generic cell assumes the form

$$\begin{aligned} n_{i,j,t+1} = {} & p_{i-1,i}\, n_{i-1,j,t} + p_{i+1,i}\, n_{i+1,j,t} + p_{i,i}\, n_{i,j,t} + p_{j-1,j}\, n_{i,j-1,t} \\ & + p_{j+1,j}\, n_{i,j+1,t}, \end{aligned} \tag{A2.7}$$

with $n_{i,j,t}$ representing the number of singularities in the cell $C_{i,j}$ at time t.

According to the concrete physical situation the transition probabilities assume a fixed form allowing consequently to precise the structure of the master stochastic equation.

A2.1. Homogeneous Medium

In the case of Snell's refraction law and the Fermat's principle of minimum time we had two optical different mediums. In such a case we can study the behaviour of the master stochastic equation in a homogeneous medium.

The eurhythmic probabilities of transition in a homogeneous medium, where $I_{i,j}$ = constant, are, of course, also constant, as can be seen

$$\tilde{p}_{i,i-1} = \frac{I_{i-1,j}}{I_{i-1,j} + I_{i+1,j} + I_{i,j} + I_{i,j-1} + I_{i,j+1}} = p = \text{constant.} \tag{A2.8}$$

Assuming that the components due to the initial velocity are such that

$$q_{oh} = q_{ov} = 0, \tag{A2.9}$$

In such conditions the probabilities of transition can be written

$$p_{i-1,i} = p_h, \quad p_{i+1,i} = q_{h,} \quad p_{i,i} = \delta, \quad p_{j-1,j} = p_v, \quad p_{j+1,j} = q_v, \tag{A2.10}$$

with

$$p_h = p + p_{oh}, \quad q_h = p, \quad \delta = p, \quad p_v = p + p_{ov}, \quad q_v = p, \tag{A2.11}$$

which by substitution in the stochastic master equation (A2.7) gives

$$n_{i,j,t+1} = p_h\, n_{i-1,j,t} + q_h\, n_{i+1,j,t} + \delta\, n_{i,j,t} + p_v\, n_{i,j-1,t} + q_v\, n_{i,j+1,t}. \tag{A2.12}$$

This discrete equation can be approached to a continuous equation of the form

$$\begin{aligned} n(x,y,t+\tau_0) = {} & p_h\, n(x-\ell_0,y,t) + q_h\, n(x+\ell_0,y,t) + \delta\, n(x,y,t) \\ & + p_v\, n(x,y-\ell_0,t) + q_v\, n(x,y+\ell_0,t) \end{aligned} \tag{A2.13}$$

that can be expanded in Taylor series giving

$$\begin{aligned} & n(x,y,t) + \tau_0 \frac{\partial n(x,y,t)}{\partial t} + \frac{1}{2!}\tau_0^2 \frac{\partial^2 n(x,y,t)}{\partial t^2} + \ldots \\ & = p_h n(x,y,t) - p_h \ell_0 \frac{\partial n(x,y,t)}{\partial x} + \frac{1}{2!} p_h \ell_0^2 \frac{\partial^2 n(x,y,t)}{\partial x^2} - \frac{1}{3!} p_h \ell_0^3 \frac{\partial^3 n(x,y,t)}{\partial x^3} + \ldots \\ & \quad + q_h n(x,y,t) + q_h \ell_0 \frac{\partial n(x,y,t)}{\partial x} + \frac{1}{2!} q_h \ell_0^2 \frac{\partial^2 n(x,y,t)}{\partial x^2} + \frac{1}{3!} q_h \ell_0^3 \frac{\partial^3 n(x,y,t)}{\partial x^3} + \ldots \\ & \quad + \delta\, n(x,y,t) \\ & \quad + p_v n(x,y,t) - p_v \ell_0 \frac{\partial n(x,y,t)}{\partial y} + \frac{1}{2!} p_v \ell_0^2 \frac{\partial^2 n(x,y,t)}{\partial y^2} - \frac{1}{3!} p_v \ell_0^3 \frac{\partial^3 n(x,y,t)}{\partial y^3} + \ldots \\ & \quad + q_v n(x,y,t) + q_v \ell_0 \frac{\partial n(x,y,t)}{\partial y} + \frac{1}{2!} q_v \ell_0^2 \frac{\partial^2 n(x,y,t)}{\partial y^2} + \frac{1}{3!} q_v \ell_0^3 \frac{\partial^3 n(x,y,t)}{\partial y^3} + \ldots \end{aligned} \tag{A2.14}$$

since the linear size ℓ_0 of the cell is very small and furthermore the velocity of transition is very big implying that

$$\ell_0 >>> t_0 \equiv \tau_0 \tag{A2.15}$$

it is reasonable to cut the expression (A2.14) from τ_0^2 and ℓ_0^3 giving

$$\begin{aligned} n + \tau_0 \frac{\partial n}{\partial t} = {} & (p_h + q_h + \delta + p_v + q_v)\, n \\ & - \ell_0 (p_h - q_h) \frac{\partial n}{\partial x} + \frac{1}{2} \ell_0^2 (p_h + q_h) \frac{\partial^2 n}{\partial x^2} \\ & - \ell_0 (p_v - q_v) \frac{\partial n}{\partial y} + \frac{1}{2} \ell_0^2 (p_v + q_v) \frac{\partial^2 n}{\partial y^2} \end{aligned} \tag{A2.16}$$

recalling that

$$p_h + q_h + \delta + p_v + q_v = 1,$$

and making

$$\mu_h = p_h - q_h = p_{oh}, \quad \mu_v = p_v - q_v = p_{ov}, \tag{A2.17}$$

$$D_h = p_h + q_h = 2p + p_{oh}, \quad D_v = p_v + q_v = 2p + p_{ov} \tag{A2.18}$$

which are the drift μ components and the diffusion coefficients D, along the horizontal and vertical axes. In this situation Eq. A2.16 can be written in the condensed form

$$D_h \frac{\partial^2 n}{\partial x^2} + D_v \frac{\partial^2 n}{\partial y^2} - \frac{2\mu_h}{\ell_0} \frac{\partial n}{\partial x} - \frac{2\mu_v}{\ell_0} \frac{\partial n}{\partial y} = \frac{2}{\ell_0 v_0} \frac{\partial n}{\partial t}. \tag{A2.19}$$

A2.1.1. The Drift Describes a Preferential Motion in a Stochastic Process

In order to see that the drift indeed describes a preferential sense for the motion in a stochastic process we are going to write the two-dimensional stationary stochastic equation for the case when the drift has only components along the horizontal axis.

In this situation looking at Fig. A2.1, and recalling that the number of corpuscles in cell (i,j) is given by the number of those that remain plus the ones that enter the cell, we are led to write the balance equation

$$n_{i,j} = p_h n_{i-1,j} + q_h n_{i+1,j} + \delta\, n_{i,j} + p_v n_{i,j-1} + q_v n_{i,j+1}, \tag{A2.20}$$

which can also be written

$$
\begin{aligned}
n(x,y) = p_h n(x-\ell_0,y) + q_h n(x+\ell_0,y) + \delta\, n(x,y) + p_v n(x,y-\ell_0) \\
+ q_v n(x,y+\ell_0).
\end{aligned} \tag{A2.21}
$$

This equation can be developed in Taylor power series giving after the habitual cutoff processes

$$
\begin{aligned}
n = {} & p_h n - \ell_0 p_h \frac{\partial n}{\partial x} + \frac{1}{2}\ell_0^2 p_h \frac{\partial^2 n}{\partial x^2} \\
& + q_h n + \ell_0 q_h \frac{\partial n}{\partial x} + \frac{1}{2}\ell_0^2 q_h \frac{\partial^2 n}{\partial x^2} \\
& + \delta\, n \\
& + p_v n - \ell_0 p_v \frac{\partial n}{\partial y} + \frac{1}{2}\ell_0^2 p_v \frac{\partial^2 n}{\partial y^2} \\
& + q_v n + \ell_0 q_v \frac{\partial n}{\partial y} + \frac{1}{2}\ell_0^2 q_v \frac{\partial^2 n}{\partial y^2}
\end{aligned} \tag{A2.22}
$$

or

$$
\frac{1}{2}\ell_0^2(p_h+q_h)\frac{\partial^2 n}{\partial x^2} + \frac{1}{2}\ell_0^2(p_v+q_v)\frac{\partial^2 n}{\partial y^2} - \ell_0(p_h-q_h)\frac{\partial n}{\partial x} - \ell_0(p_v-q_v)\frac{\partial n}{\partial y} = 0
$$

giving

$$
D_h\frac{\partial^2 n}{\partial x^2} + D_v\frac{\partial^2 n}{\partial y^2} - \frac{2\mu_h}{\ell_0}\frac{\partial n}{\partial x} - \frac{2\mu_v}{\ell_0}\frac{\partial n}{\partial y} = 0, \tag{A2.23}
$$

expression that, as expected, is equal to (A2.19) when the number of corpuscles does not depend explicitly on time.

Since we have assumed that the drift was defined only along the horizontal axis,

$$
\mu_h = \mu, \quad \mu_v = 0 \tag{A2.24}
$$

we have finally

$$
D_h\frac{\partial^2 n}{\partial x^2} + D_v\frac{\partial^2 n}{\partial y^2} - \frac{2\mu}{\ell_0}\frac{\partial n}{\partial x} = 0. \tag{A2.25}
$$

A solution to this differential equation has the form

$$
n(x,y) = A\, e^{(k_1 x + i k_1' x) - k_2|y|} + B = \phi + B, \tag{A2.26}
$$

as can easily be seen. In fact:

$$\frac{\partial n}{\partial x} = (k_1 + i k_1^{'})\phi, \quad \frac{\partial^2 n}{\partial x^2} = (k_1^2 - k_1^{'2} + 2i k_1 k_1^{'})\phi, \quad \frac{\partial^2 n}{\partial y^2} = k_2^2 \phi, \tag{A2.27}$$

which by substitution in (A2.25) gives after separating the real and imaginary parts

$$\begin{cases} D_x(k_1^2 - k_1^{'2}) + D_y k_2^2 - \dfrac{2\mu}{\ell_0} k_1 = 0 \\ D_x k_1 - \dfrac{\mu}{\ell_0} = 0 \end{cases} \tag{A2.28}$$

giving

$$k_1 = \frac{\mu}{\ell_0 D_x}, \tag{A2.29}$$

where we have renamed

$$D_x = D_h, \quad D_y = D_v. \tag{A2.30}$$

Now a possible relation that much simplifies the result comes from making

$$k_1^{'} = k_1 \tag{A2.31}$$

in such conditions the first equation of (A2.28) transforms into

$$D_y k_2^2 - \frac{2\mu}{\ell_0}\frac{\mu}{\ell_0 D_x} = 0$$

giving

$$k_2 = \sqrt{\frac{2}{D_x D_y}}\frac{\mu}{\ell_0}, \tag{A2.32}$$

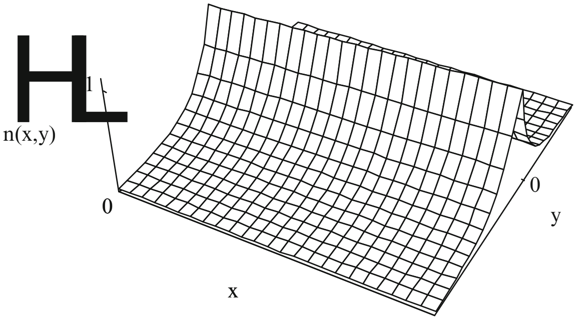
consequently a solution to the balance equation (A2.25) can then be written explicitly

$$n(x, y) = A\, e^{\frac{\mu}{\ell_0 D_x} x + i \frac{\mu}{\ell_0 D_x} x - \frac{\sqrt{2}\mu}{\ell_0 \sqrt{D_x D_y}} |y|} + B, \tag{A2.33}$$

which overall graphic representation of the real part is seen in Fig. A2.2

Needless to say that the drift $\mu \geqslant 0$ must be very small otherwise there would be a practically complete deterministic motion. On the other hand the $\vec{k}$ wave vector is

Fig. A2.2 Plot of the real part of the solution stationary stochastic equation



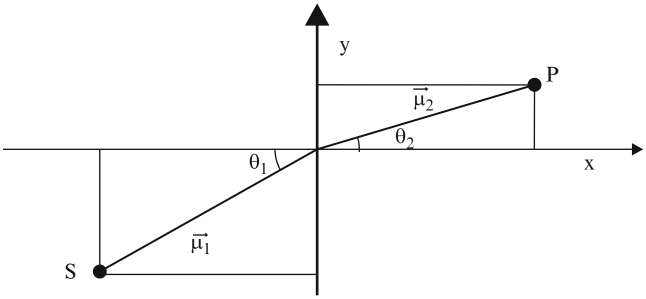


Fig. A2.3 Average motion of the corpuscles in two optical mediums

only along the xx direction, which means, as we have said before, that the drift for all average purposes characterizes indeed the preferential direction of the corpuscles, meaning that

$$\vec{\mu} \propto \vec{k}. \tag{A2.34}$$

In such circumstances we can draw the graphic representing average motion of the corpuscles in two homogeneous mediumswhere the drift has the form

$$\vec{\mu} = \mu_x \vec{e}_x + \mu_y \vec{e}_y. \tag{A2.34$'$}$$

From Fig. A2.3 it is seen that

$$\begin{cases} \mu_{1y} = \mu_1 \sin \theta_1 \\ \mu_{2y} = \mu_2 \sin \theta_2 \end{cases} \tag{A2.35}$$

because along the yy direction there is no change in the medium we are allowed to make

$$\mu_1 \sin\theta_1 = \mu_2 \sin\theta_2. \tag{A2.36}$$

Since by (A2.34)

$$\mu_x = \alpha\, k_x; \quad \mu_y = \alpha\, k_y, \tag{A2.37}$$

where α is a proportionality constant we also have

$$k_1 \sin\theta_1 = k_2 \sin\theta_2, \tag{A2.38-1}$$

recalling that $k = 2\pi/\lambda$, we have by substitution

$$\frac{2\pi}{\lambda_1} \sin\theta_1 = \frac{2\pi}{\lambda_2} \sin\theta_2, \tag{A2.38-2}$$

or

$$\frac{T}{\lambda_1} \sin\theta_1 = \frac{T}{\lambda_2} \sin\theta_2, \tag{A2.38-3}$$

because in a linear optical mediums $T_1 = T_2$, and $v = \lambda/T$ we can write

$$\frac{1}{v_1} \sin\theta_1 = \frac{1}{v_2} \sin\theta_2, \tag{A2.38-4}$$

which multiplied by c gives finally Snell's formula

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \tag{A2.38-5}$$

This formula means, according to Fermat, that light, that is the corpuscles of light, follow a path such that the time taken in the whole course from S to P is minimum.

A3. Derivation of Galilee's Law from the Principle of Eurhythmy

For this derivation we assume that the theta wave intensity gravitic field has a linear variation increasing when approaching the Earth. Furthermore, it is also understood that the stochastic model can be simplified from two dimensions to one single dimension which much simplifies the calculations.

A3.1. One-dimensional Approach

The probabilities of transition constant in time A3.1 are seen in Fig. A3.1

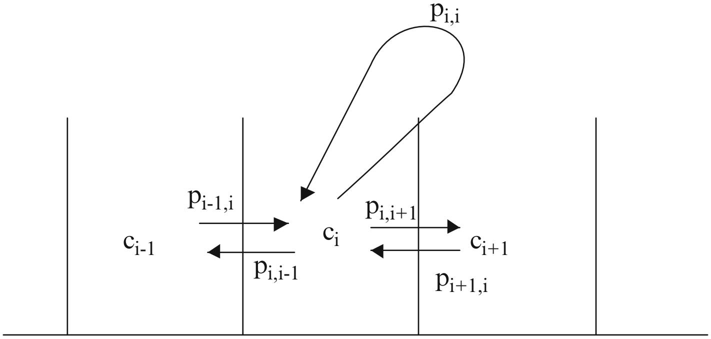


Fig. A3.1 Transition probabilities for the generic cell in one-dimensional field

with

$$q_i = p_{i,i-1}; \quad \delta_i = p_{i,i}; \qquad p_i = p_{i,i+1} \tag{A3.1}$$

and

$$q_i + \delta_i + p_i = 1. \tag{A3.2}$$

These transition probabilities can be expressed in terms of the intensity according to the principle of eurhythmy

$$\begin{cases} q_i = p_{i.i-1} = \dfrac{I_{i-1}}{I_{i-1} + I_i + I_{i+1}} \\ \delta_i = p_{i.i-1} = \dfrac{I_i}{I_{i-1} + I_i + I_{i+1}} \\ p_i = p_{i.i-1} = \dfrac{I_{i+1}}{I_{i-1} + I_i + I_{i+1}} \end{cases} \tag{A3.3}$$

The stochastic evolution equation can be written recalling that the number of corpuscles in the generic cell C_i in the instant *t +1* is equal to the number of corpuscles that remain plus the ones that enter from the right and the left

$$n_{i,t+1} = p_{i-1} n_{i-1,t} + \delta_i n_{i,t} + q_{i+1} n_{i+1,t}, \tag{A3.4}$$

or

$$n_{i,t+1} = (1 - q_i - p_i)n_{i,t} + p_{i-1}n_{i-1,t} + q_{i+1}n_{i+1,t} \tag{A3.4$'$}$$

$$n_{i,t+1} = n_{i,t} - q_i n_{i,t} - p_i n_{i,t} + p_{i-1}n_{i-1,t} + q_{i+1}n_{i+1,t}. \tag{A3.4$''$}$$

Naming

$$\begin{cases} t_i = p_i n_i \\ r_i = q_i n_i \end{cases} \tag{A3.5}$$

the stochastic equation assumes the form

$$n_{i,t+1} = n_{i,t} - r_{i,t} - t_{i,t} + t_{i-1,t} + r_{i+1,t}, \tag{A3.6}$$

which can be approached to the continuous form

$$n(x, t + \tau_0) = n(x, t) - r(x, t) - t(x, t) + t(x - \ell_0, t) + r(x + \ell_0, t). \tag{A3.7}$$

This formula can be Taylor expanded, with the customary cutoff criteria giving

$$\begin{aligned} n + \tau_0\, n_t &= n - r - t + t - \ell_0\, t_x + \frac{1}{2}\ell_0^2\, t_{xx} + r + \ell_0\, r_x + \frac{1}{2}\ell_0^2\, r_{xx} \\ \tau_0\, n_t &= -\ell_0\, t_x + \frac{1}{2}\ell_0^2\, t_{xx} + \ell_0\, r_x + \frac{1}{2}\ell_0^2\, r_{xx}, \end{aligned} \tag{A3.8}$$

where

$$n_t = \frac{\partial n}{\partial t}, \quad t_x = \frac{\partial t}{\partial x}, \quad t_{xx} = \frac{\partial^2 t}{\partial x^2}, \quad \cdots \tag{A3.9}$$

but

$$\begin{cases} t = p\, n \\ t_x = p_x n + p\, n_x \\ t_{xx} = p_{xx} n + 2 p_x n_x + p\, n_{xx} \end{cases} \qquad \begin{cases} r = q\, n \\ r_x = q_x n + q\, n_x \\ r_{xx} = q_{xx} n + 2 q_x n_x + q\, n_{xx} \end{cases} \tag{A3.10}$$

then by substitution in (A3.8) we got

$$\begin{aligned} \tau_0\, n_t =& - \ell_0\, p_x\, n - \ell_0\, p\, n_x + \frac{1}{2}\ell_0^2\, p_{xx}\, n + \ell_0^2\, p_x\, n_x + \frac{1}{2}\ell_0^2\, p\, n_{xx} \\ & + \ell_0\, q_x\, n + \ell_0\, q\, n_x + \frac{1}{2}\ell_0^2\, q_{xx}\, n + \ell_0^2\, q_x\, n_x + \frac{1}{2}\ell_0^2\, q n_{xx} \end{aligned} \tag{A3.11}$$

or

$$\begin{aligned}\tau_0\, n_t =& -\ell_0 (p_x - q_x)\, n - \ell_0\, (p-q)\, n_x + \frac{1}{2}\ell_0^2\, (p_{xx} + q_{xx})\, n \\ & + \ell_0^2\, (p_x + q_x)\, n_x + \frac{1}{2}\ell_0^2\, (p+q)\, n_{xx}\end{aligned} \tag{A3.11$'$}$$

since

$$\begin{cases} p - q = \mu \\ p_x - q_x = \mu_x \end{cases} \qquad \begin{cases} p + q = D \\ p_x + q_x = D_x \\ p_{xx} + q_{xx} = D_{xx} \end{cases} \tag{A3.12}$$

we have

$$\frac{\tau_0}{\ell_0}\, n_t = \left(\frac{1}{2}\ell_0\, D_{xx} - \mu_x\right) n + (\ell_0\, D_x - \mu)\, n_x + \frac{1}{2}\ell_0\, D n_{xx}, \tag{A3.13}$$

making

$$A(x) = \frac{1}{2}\ell_0 D, \quad B(x) = \ell_0 D_x - \mu, \quad C(x) = \frac{1}{2}\ell_0 D_{xx} - \mu_x, \tag{A3.14}$$

and by substitution in (A3.13) it gives the generic form for the fundamental stochastic evolution equation

$$A(x)\frac{\partial^2 n}{\partial x^2} + B(x)\frac{\partial n}{\partial x} + C(x)\, n = \frac{1}{v_0}\frac{\partial n}{\partial t}. \tag{A3.15}$$

The explicit form for the transition probabilities can be obtained from Eqs. A3.3 which may be written

$$\begin{cases} q(x) = \dfrac{I(x-\ell_0)}{I(x-\ell_0) + I(x) + I(x+\ell_0)} \\[2ex] \delta(x) = \dfrac{I(x)}{I(x-\ell_0) + I(x) + I(x+\ell_0)} \\[2ex] p(x) = \dfrac{I(x+\ell_0)}{I(x-\ell_0) + I(x) + I(x+\ell_0)} \end{cases} \tag{A3.16}$$

when using the integration rectangle formula.

Expanding in Taylor power series and making the cutoff at the habitual positions we have

$$I(x \pm \ell_0) \cong I \pm \ell_0\, I_x + \frac{1}{2}\ell_0^2\, I_{xx}$$

$$\begin{aligned} I(x-\ell_0) + I(x) + I(x+\ell_0) &= I - \ell_0\, I_x + \frac{1}{2}\ell_0^2\, I_{xx} + I + I + \ell_0\, I_x + \frac{1}{2}\ell_0^2\, I_{xx} \\ &= 3I + \ell_0^2\, I_{xx} \end{aligned} \tag{A3.17}$$

that by substitution in (A3.16) gives

$$\begin{cases} q(x) = \dfrac{I - \ell_0\, I_x + \frac{1}{2}\ell_0^2\, I_{xx}}{3I + \ell_0^2 I_{xx}} \\ \delta(x) = \dfrac{I}{3I + \ell_0^2 I_{xx}} \\ p(x) = \dfrac{I + \ell_0\, I_x + \frac{1}{2}\ell_0^2\, I_{xx}}{3I + \ell_0^2 I_{xx}} \end{cases} \tag{A3.18}$$

A3.2. *Linear Approximation*

Now following the initial approximation that the intensity of the theta wave field as a linear variation in the domain of validity of Galilee's law

$$I(x) = \alpha\, x \tag{A3.19}$$

we have by substitution in (A3.18)

$$\begin{cases} q(x) = \dfrac{x - \ell_0}{3x} \\ \delta(x) = \dfrac{1}{3} \\ p(x) = \dfrac{x + \ell_0}{3x} \end{cases} \tag{A3.20}$$

From the concrete forms for the transition probabilities it is possible to obtain directly the drift and diffusion coefficients and its derivatives

$$\begin{cases} \mu(x) = p(x) - q(x) = \dfrac{2}{3}\dfrac{\ell_0}{x} \\ D(x) = p(x) + q(x) = \dfrac{2}{3} \end{cases} \qquad \begin{cases} \mu_x = -\dfrac{2}{3}\dfrac{\ell_0}{x^2} \\ D_x = D_{xx} = 0 \end{cases} \tag{A3.21}$$

which by substitution in Eq. A3.14 gives for the coefficients

$$A(x) = \frac{1}{3}\ell_0, \quad B(x) = -\frac{2}{3}\frac{\ell_0}{x}, \quad C(x) = \frac{2}{3}\frac{\ell_0}{x^2}, \tag{A3.22}$$

and by substitution in Eq. A3.15 we finally obtain for the master stochastic equation for the process

$$\frac{\partial^2 n}{\partial x^2} - \frac{2}{x}\frac{\partial n}{\partial x} + \frac{2}{x^2} n = \frac{3}{\ell_0 v_0}\frac{\partial n}{\partial t}. \tag{A3.23}$$

In order to integrate this equation it is convenient to make the transformation

$$n = x\,\phi. \tag{A3.24}$$

In this case the derivatives give

$$\begin{cases} \dfrac{\partial n}{\partial x} = \dfrac{\partial}{\partial x}(x\phi) = \phi + x\dfrac{\partial \phi}{\partial x} \\ \dfrac{\partial^2 n}{\partial x^2} = 2\dfrac{\partial \phi}{\partial x} + x\dfrac{\partial^2 \phi}{\partial x^2} \\ \dfrac{\partial \phi}{\partial x} = \dfrac{\partial}{\partial x}\left(\dfrac{1}{x}n\right) = -\dfrac{1}{x^2}n + \dfrac{1}{x}\dfrac{\partial n}{\partial x} = -\dfrac{1}{2}\left(-\dfrac{2}{x}\dfrac{\partial n}{\partial x} + \dfrac{2}{x^2}n\right) \end{cases} \tag{A3.25}$$

and inserting in (A3.23)

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{3}{\ell_0 v_0}\frac{\partial \phi}{\partial t}. \tag{A3.26}$$

A solution to this equation is

$$\phi = Ae^{-k\,x+\omega\,t} + Be^{+k\,x+\omega\,t}, \quad \omega = \frac{k^2\ell_0 v_0}{3}, \tag{A3.27}$$

then the solution for the master stochastic equation assumes the form

$$n = Axe^{-k\,x+\omega\,t} + Bxe^{+k\,x+\omega\,t}. \tag{A3.28}$$

In order to fix the value of the constants we assume that at the initial time the number of corpuscles in the starting cell is

$$n(\ell_0, 0) = n_0 \tag{A3.29}$$

so

$$n(\ell_0, 0) = n_0 = A\,\ell_0 e^{-k\ell_0} + B\,\ell_0 e^{+k\ell_0}. \tag{A3.30}$$

On the other side it is necessary to determinate the average time that a corpuscle takes to reach the generic cell S, such that $S\,\ell_0 = \ell$, after traveling the path of length ℓ. This situation occurs when

$$n(\ell,t) = n_0, \quad \text{and} \quad n_0 >>> 1. \tag{A3.31}$$

This condition implies that

$$A\,\ell e^{-k\ell+\omega t} + B\,\ell e^{+k\ell+\omega t} = A\,\ell_0 e^{-k\ell_0} + B\,\ell_0 e^{+k\ell_0}. \tag{A3.32}$$

Now the value of the constants need to be such that they contain Galilee's law stating that the traveled spaces are proportional to the square of the times taken

$$\ell \propto t^2, \quad t = \beta\,\ell^{\frac{1}{2}}. \tag{A3.33}$$

In such conditions by substitution in (A3.32) we have

$$A\,\ell e^{-k\ell+\omega\,\beta\,\ell^{\frac{1}{2}}} + B\,\ell e^{+k\ell+\omega\,\beta\,\ell^{\frac{1}{2}}} = A\,\ell_0 e^{-k\ell_0} + B\,\ell_0 e^{+k\ell_0}$$

then

$$B = \frac{\ell_0 e^{-k\ell_0} - \ell e^{-k\ell+\omega\,\beta\,\ell^{\frac{1}{2}}}}{-\ell_0 e^{+k\ell_0} + \ell e^{+k\ell+\omega\,\beta\,\ell^{\frac{1}{2}}}} A\,. \tag{A3.34}$$

Now assuming the relationship (A3.34) between the two constants for the solution (A3.28)

$$n = Axe^{-k\,x+\omega\,t} + A\frac{\ell_0 e^{-k\ell_0} - \ell\,e^{-k\,\ell+\omega\,\beta\,\ell^{\frac{1}{2}}}}{-\ell_0 e^{+k\ell_0} + \ell\,e^{+k\,\ell+\omega\,\beta\,\ell^{\frac{1}{2}}}}\,x\,e^{+k\,x+\omega\,t} \tag{A3.35}$$

and by an inverse reasoning we can obtain, directly Galilee's law

$$\ell \propto t^2.$$

Acknowledgements I want to thank Professors: R.N. Moreira and Gildo Magalhães and Dr. J. Cordovil for their encouragement and support in the development of the ideas that have given rise to the present work. I also want to thank the referee for his pertinent comments that much help to improve the final form of the manuscript. This work was supported by FCT.

References

Chui, K. 1992. *An introduction to wavelets*. New York: Academic.

Croca, J.R. 2003. *Towards a nonlinear quantum physics*. London: World Scientific.

Croca, J.R. 2007. *Local wavelet analysis versus nonlocal Fourier analysis.* Accepted for publication in *International Journal of Quantum Information* Vol.5, Nos. 1&2, 1–7.

de Broglie, L. 1964. *The current interpretation of the wave mechanics, a critical study*. Amsterdam: Elsevier.

de Broglie, L., and J.L. Andrade e Silva. 1971. *La Réinterprétation de la mecanique ondulatoire, Tome 1 Principes Géneraux*. Paris: Gauthier-Villards.

Eckt, E., and S. Zajac. 1974. *Optics*. Boston: Addison-Wesley.

Edwards, Matthew R. 2002. *Le Sage's theory of gravitation. The revival by Kelvin and some later developments, in pushing gravity, new perspectives on Le Sage's theory of gravitation*. Montreal: Apeiron.

Evans, James. 2002. *Gravity in the century of light, in pushing gravity, new perspectives on Le Sage's theory of gravitation*. Montreal: Apeiron.

Hubbard, B.B. 1998. *The word according to wavelets*. Natick: A.K. Peters Wellesley.

Rica da Silva, J.R. Croca and R.N. Moreira, 1998. *Non-Linear Schrödinger Equation, Burger's Equation and Superposition of Solutions, in Causality and Locality in Modern Physics and Astronomy,* eds. S. Jeffers et al. (Kluwer Academic Publishers).

Chapter 4
Unifying Science Through Computation: Reflections on Computability and Physics

Edwin J. Beggs, José Félix Costa, and John V. Tucker

Abstract Many activities of a contemporary working scientist involve the idea of the unity of science. There are countless examples where the ideas and methods of one subject find application in another. There are subjects that comfortably straddle the border between disciplines. There are problems that can only be tackled by multidisciplinary approaches. Science is a loose federation of diverse intellectual, experimental and material communities and cultures. However, these cultures are strong.

In this paper we reflect upon an area of research that is attracting the attention of computer scientists, mathematicians, physicists and philosophers: the relationship between theories of computation and physical systems. There are intriguing questions about the computability of physics, and the physical foundations of computability, that can set the agenda for a new subject, and that will not go away. Research is in an early phase of its development, but has considerable potential and ambition.

First, we will argue that concepts of computability theory have a natural place in physical descriptions. We will look at incomputability and (1) the idea

---

E.J. Beggs (✉) • J.V. Tucker
School of Physical Sciences, Swansea University, Singleton Park, Swansea, SA3 2HN, Wales, United Kingdom
e-mail: E.J.Beggs@swansea.ac.uk; J.V.Tucker@swansea.ac.uk

J.F. Costa
School of Physical Sciences, Swansea University, Singleton Park, Swansea, SA3 2HN, Wales, United Kingdom

Department of Mathematics, Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, Portugal

Centro de Matemática e Aplicações Fundamentais do Complexo Interdisciplinar, Universidade de Lisboa, Lisbon, Portugal
e-mail: fgc@math.ist.utl.pt

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology, and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_4,

that computers “exist” in Nature, (2) models of physical systems and notions of prediction, and (3) hypercomputation. We will reflect upon computability and physics as an example of work crossing the frontiers of two disciplines, introducing new questions and ways of argument in physics, and enabling a reappraisal of computers and computation. We will also notice the social phenomenon of suspicion and resistance, as the theories are unbalanced by their encounter with one another.

Keywords Computability • Hypercomputation • Nature • Oracle • Predictability

4.1 Introduction

Scientists are surrounded by references to the unity of science. They can be found in tales about the historical development of science, and in the theories and practices of contemporary science. Long ago, biology was invaded by chemistry; chemistry was invaded by physics; and, long before, physics was conquered by mathematics. References can also be found in public policies about science—some sort of unity must be assumed to make sense of the ever fashionable desire for interdisciplinary and multidisciplinary research, for example. New subjects are born of old, such as computer science of mathematics, electronics, logic and linguistics. Computer science is trying to invade everything.

Historically, there are plenty of examples where the ideas and methods of one subject find application to problems in another. In judging the application, the “distance” between the subjects involved and the scope for new developments from the application are important criteria: the further apart, the more remarkable; the larger the legacy, the more significant. Perhaps, a whole new subject is formed that straddles the border between disciplines.

Mathematicians and physicists have a deep faith in the unity of their disciplines and make use of this unity in their research. They are blessed with long memories and long term goals. There are extraordinary examples in mathematics and physics. In mathematics, the twentieth century saw algebra and topology combine with dramatic effects. Poincaré’s fundamental group of a space is indeed a beautiful innovation but it is an amazingly humble origin for diverse new mathematics, including concrete topics, such as *combinatorial group theory* and *knot theory* to grandiose theories of everything, such as *category theory*. In physics, there is the study of conservation of energy from Grove to Helmholtz, or the on-going search to unify quantum theory and relativity for a physical theory of everything. Between mathematics and physics, the emergence of non-Euclidean geometry and its subsequent role in relativity is an example of unity. Mathematics and physics export great deal to other disciplines, but it all takes a very long time.

Computation is a unifying force in science: computers and software are *everywhere*. Why? Quantification, a fundamental process of science, rests upon the collection, generation, storage, processing and interpretation of data. *Therefore*, technologies for data have long been essential. Computer Science is the new discipline whose core concepts are data and algorithm. Actually, because of quantification, *the concepts of data and algorithm can be found everywhere*. What do our fundamental *theories* of data and algorithms have to offer science?

Computability theory, founded by Church, Turing and Kleene in 1936, is a deep theory for the functions computable by algorithms on particular forms of finite discrete data, such as strings over $\{0, 1\}$ and natural numbers $\{0, 1, 2, \ldots\}$. Digital data is *precisely* the data that can have finite representations, coded by strings or naturals. Computability theory has been extended to arbitrary data via generalisations to abstract algebras (Stoltenberg-Hansen and Tucker 1995; Tucker and Zucker 2000), and, in particular, to continuous data, such as real numbers, via approximations (Tucker and Zucker 2000, 2004). Computability theory is at the heart of our understanding of data and algorithm. What has computability theory to offer scientific understanding?

In this paper we reflect upon the relationship between *theories* of computability and physical systems. There are intriguing questions about the computability of physics, and the physical foundations of computability, that can set the agenda for a new subject—they are questions that will not go away. We have a great deal knowledge to call upon. It is an area of research that is attracting the attention of computer scientists, mathematicians, physicists and philosophers. Research is in an early phase of its development, but has considerable potential and ambition.

First, we will argue that concepts from computability theory have a natural place in physical descriptions: we show how abstract machine models with oracles can frame models of the solar system. A connection between computability and physics introduces a connection between *in*computability and physics. We discuss three "causes" of incomputability:

1. *Partial or insufficient information for computation*;
2. *Unpredictability of properties of a model*;
3. *Hypercomputational phenomena in the Universe*.

We reflect upon computability and physics as an example of work at the frontiers of two disciplines. It includes the introduction of new questions and ways of reasoning in physics, and enables a re-appraisal of what makes a computer. We also notice the interesting social phenomenon of suspicion and resistance, as the theories are knocked off balance by bumping into one another.

There is a great deal of background to this theoretical task. Computability Theory is being redeveloped, even reborn. The mathematical subject that was created by philosophical problems in the foundations of mathematics in the 1930s has seen fantastic technical developments and spectacular applications since then. One need only ponder Rogers text-book (Rogers 1968) of 1968(?). But, for a period of at least 20 years, roughly the 1970–1990s, its intellectual vibrancy and centrality has been eclipsed by its intense technical development. Technicians forgot or avoided old messy debates about what computability theory is actually about, and whether it has useful consequences for mathematics and science generally. The image of computability theory was dominated by its internal technical agenda and achievements, perhaps most extremely expressed by *generalized recursion theory* (see Cooper and Odifreddi 2003; Fenstad 1980, 2001). For many computability theorists thinking about old debates, unfinished business, new applications and the education of young scientists, the prevailing technocratic view was to become an irritating problem. In the broad community of researchers in mathematical logic, theoretical computer science and philosophy, computability theory was considered a corpse for pure mathematicians to dissect.

The rebirth of computability theory involves a large scale investigation of fundamental questions of about computation. Are physical systems computable? What can computers, based upon the new technologies of quantum information, optics, etc., compute? These questions come from outside computability theory and confront it with questions that will not go away and uncomfortable notions like hypercomputation—can the technologies compute more, or more efficiently, than Turing machines? From inside computability theory, Cooper and Odifreddi have pressed for the exploration of how Turing's universe lies embedded in Nature (Cooper and Odifreddi 2003).

Of course, the classical mathematical theory had long been at home with the non-computable; for example, through various kinds of hierarchies (arithmetic, hyperarithmetic, and analytic (Hinman 1978)) and, of course, the study of non-computable Turing degrees such as $0', 0'', \ldots,$ to say nothing of the extremes of generalisations of computability theory to ordinals, searching for priority arguments to solve analogues of Post's Problem.

However, the new approach to the foundations of computability leads to fundamental questions about what are non-computable sets and functions; these lead to debate and controversy: "hypercomputation" becomes a forbidden concept because it corresponds to a concept of implementable that has the potential to contradict the Church-Turing Thesis, the primary legacy of the theory. We comment briefly on the origins of this criticism and misinterpretations of concepts such as super-Turing computational power.

Our reflections here are an initial attempt to examine the wider context of our current research programmes (Beggs et al. 2007, 2008a,b; Beggs and Tucker 2008, 2004, 2006, 2007a,b).

4.2 Computability in Nature: Stonehenge as a Calculator with Oracles

Let us reflect on a seemingly complicated example of computation intimately related with Nature.

The astronomer Fred Hoyle showed in Hoyle (1972, 1977) that Stonehenge can be used to predict the solar and the lunar eclipse cycles. Now, for our purposes, it *does not matter* whether the Ancient Britons did, or did not, use this huge monument to predict the eclipse cycles; but it *does matter* that we, in our times, can use Stonehenge to make good predictions of celestial events like the azimuth of the rising Sun and of the rising Moon, or that we can use this astronomical observatory as an eclipse predictor. Hoyle's method is based upon the structure called *Stonehenge I*;

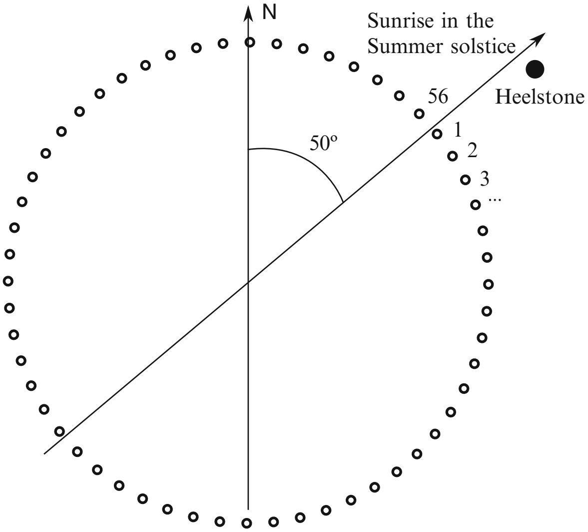


Fig. 4.1 A schematic drawing of Stonehenge I

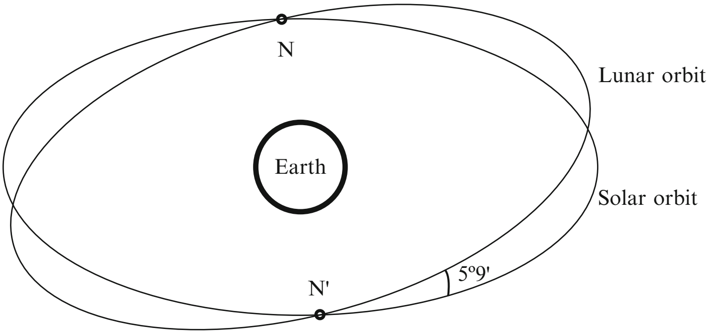


Fig. 4.2 Orbits

important in this computational task is the alignment of the Heelstone with the summer solstice, and the circle of *Aubrey holes*, made of 56 stones (see Fig. 4.1), buried until the seventeenth century, and discovered by the antiquary John Aubrey.

Hoyle's algorithm makes use of three counters for the task: the first counter, one little stone representing the sun, counts the days of the year along the circle of 56 Aubrey holes; the second counter, representing the moon, counts the days of the lunar month; finally, a third counter, takes care of the Metonic cycle, in which the same phases of the moon are repeated on the same date of the year to within an hour or so after a period of 19 years, a fact discovered by Meton around 430 B.C. though it is believed to have been known earlier. In other words the third small stone counts along the cycle of the lunar node, one of the intersection points of the ecliptic with the Moon's orbit (Fig. 4.2).

Since $56 \times \frac{13}{2} = 365$, the first counter has to move two places—two Aubrey holes—each 13 days (one place per week roughly speaking), counterclockwise. In a similar way, since $56 \div 2 = 28$, the second counter is allowed to move two places per day, counterclockwise. When the two counters meet at the same hole an eclipse becomes possible, but only if the Sun and the Moon are close to the lunar's node—intersection point of the ecliptic and the Moon's orbit. This point is represented by the third counter. Thus, the three counters have to meet at the same hole (more or less). This third little stone counts along the Metonic cycle: $56 \div 3 = 18.67$ (very close to the true value 18.61, a most strange coincidence), meaning that it has to move 3 places—3 Aubrey holes—per year, clockwise.

Thus the movement of the three stones around the circle of Aubrey holes allows us to predict the solar and the lunar eclipse cycles. The seminal paper, in which Stonehenge is given an astronomical interpretation as a predictor of eclipses, was by the archeologist Hawkins, in (1964), but the mathematical calculations were done by Hoyle, years later. See also Newham (2000) for a short introduction.

By calling Hoyle's method Hoyle's *algorithm*, we have introduced the idea of re-interpreting it using concepts from the theory of computation. What model of computation fits best Hoyle's algorithm? What have we got? The algorithm is based upon a simple machine with 3 counters and a memory so the n-counter machines[1] come to mind. In computability theory, Stonehenge I and Hoyle's algorithm shows that a 3-counter machine implements an eclipse cycle predictor using arithmetic modulo 56. Actually, we can also be more restrictive and think in terms of finite state automata; or we can be more general and involve Turing machines and register machines (which are "equivalent" to n counter machines for $n \geq 2$).

A quite straightforward algorithm is implemented by a special purpose analogue machine. It is analogue in the original sense that it is *analogous to the physical system it calculates*. When playing with the counter machine do we "see" in the movements of the three pebbles the Sun, a physical body, and the Moon, another physical body, and (the words are taken from Hoyle) a *holy spirit*, the lunar node, performing a dance in the sky, one that is projected onto the celestial sphere from earthly Wiltshire? Yes, if we have a Rite associated to the ballet of the stones.

However, this is not the whole story. Over time, the calculations with the counters loose accuracy. Once in a year, the Sun rises over the Heelstone. Some auxiliary stones (*the post holes*), to one side of the Heelstone, can be used to fine tune the counters: the site of the rising mid-summer Sun moves to the north and then back to the south, allowing us to fine tune the Sun's counter by observing from the center through the post holes its maximum azimuth. Some auxiliary stones also help to fine tune the second counter. The observations of the Sun and of the Moon operate like *oracles*. Thus, the calculating master, in the centre, can calculate the eclipse by the less accurate algorithm together with an oracle for the Sun and a second oracle for the Moon. Thus, *if the Ancient Britons were "aware" of algorithms then they were aware of algorithms with oracles!*.

---

[1] An n-counter machine is a register machine with n registers for natural numbers and simple assignments based upon successor and predecessor, and jumps based upon equality.

The curious thing is that we have a reality in the sky and an analogous reality in the big circle. The algorithm is captured by the real world in the sense that the real world embeds or realises or embodies the algorithm (c.f. Wolfram 2002).

Thesis 4.1. *When we abstract from physical entities, we can find special purpose computers existing in Nature.*

This idea of mathematics and computation residing in Nature can be seen in Galileo's work and we could be tempted to refer to the thesis as *Galileo's principle of natural computation*.

If computability is to be found in Nature then is incomputability to be found also? Cooper and Odifreddi address this idea in (2003) for these authors (in)-computability sounds more like an intrinsic limitation of knowledge about the Universe than a physical manifestation of hypercomputation; we will discuss this aspect later.

The idea of *The Universe* that is often used has a disadvantage. When used alone, without specification of the model we have in mind, it conveys the impression that there is a "true nature" of the Universe and that we may know it. A *universe* is simply a *model* of some aspect of the Universe. The word *universe* has the further advantage that it may be used freely and loosely without any need to remind ourselves constantly that The Universe is still mysterious and unknown.

Our Stonehenge model of a fragment of The Universe is not so bad and it certainly makes a memorable connection between Computability Theory and Physics.

Stonehenge can be also used as a calculator and computer for doing arithmetic and (why not) for implementing some more sophisticated algorithms. Moreover, Stonehenge *implements natural phenomena*, the complex movements of the Sun and the Moon in the sphere. Stonehenge possesses the means of consulting oracles in Nature itself. Through observations of the world, oracles here handle some incomputabilities.

We end this section with a question: Do the incomputabilities we mentioned above come out of

(1) An unpredictable behaviour of a model of Nature, or
(2) A really essential incomputability in Nature?

The latter supposes the existence of the hypercomputational character of some physical phenomenon. The most publicised example of this is what Penrose was looking for in (1989, 1994) and Penrose and Shimony (1997). But hypercomputational with respect to what models of computation?

Let us interpret predictability to be the ability to decide on whether or not a model has some property.

Moore was among the first to observe a failure of predictability. We are accustomed to think of chaos as sensitivity to initial conditions, which rules out reliable, reproducible predictions of system behaviour because of numerical precision. Moore showed that essential chaos exists in the sense that infinite precision in the initial conditions will not remove it. He showed that the collection of dynamical maps contains many instances of simulations of universal Turing machines. In other words, any property that a map can have, like being injective, or onto, or having an infinite domain, or having an infinite range, or being total, is undecidable unless it is trivial (cf. Rice's Theorem). Now, in terms of dynamical systems, these questions concern basins of attraction. These basins are, in general, non-computable, i.e., there is no hyper-algorithm that will tell us whether or not a point is in them. Recall Theorem 10 from Moore (1991):

Proposition 4.2.1 (Moore's undecidability theorem). *The following questions about discrete-time dynamic systems are undecidable.*

(a) Given a point x and an open set A, will x fall into A?
(b) Given a point x and a periodic point p, will x converge to p? Will a dense set of points converge to p?

Incomputability is a riddle. What kind of incomputability should we search for in Nature? Given our discussion, here are three possible causes:

1. *Partial or inadequate information for calculations*;
2. *Essential unpredictability of properties of models*;
3. *Hypercomputational phenomena in the Universe*.

In the first case, we cannot compute because we do not have all the necessary variables (e.g., the Adams-Leverrier discovery of Neptune; hidden variables in the Paris School of Quantum Mechanics). In the second case, we cannot decide algorithmically if properties of models of physical systems hold or not. In the third, Nature performs computations that our algorithms cannot.

Articles like Cooper and Odifreddi (2003) raise questions about the physical nature of computability and about the possibility of exporting new concepts from Computer Science to other sciences.

4.3 Computability in Nature: *n*-Clock Machine

Let us return to the re-interpretation of Stonehenge's counters using computability theory. The 3-counter machine has a very well known property.

Proposition 4.3.1 (Universality of *n*-Counter machine). *There is a Turing universal* 2*-counter machine.*

The n-counter machines are primitive and troublesome to program. Students play with such machines to get acquainted with a model of computation, doing exercises for calculating the sum, product, etc. How remarkable, then, that the Stonehenge computer, a 2-counter machine, computes the cycle so neatly. Thus, in the Stonehenge case, the counter machine is a kind of natural computer.

Let us consider a more sophisticated machine, one that is not well known but also fits with Stonehenge. The model was introduced in Killian and Siegelmann (1986).

Although introduced for rather ad hoc purposes in a proof, the model is attractive when reflecting on natural computing, such as in Stonehenge.

The model of computation is called the *n-alarm clock machine*, or just n-clock machine and is made of abstract clocks with programmable alarms. Time is an abstraction that belongs in the physical world. By an *alarm* in the physical world we can consider things in macroscopic or microscopic worlds that signal special events, e.g., an astronomical ephemeris, like a conjunction of planets, or the reaching of a perihelion, or an eclipse.

A n-alarm clock machine $\mathcal{A}$ consists of n clocks. Each clock $1 \leq i \leq n$ is represented by a pair (p_i, t_i), where $p_i \in \mathbb{N}$ is the period of clock i and $t_i \in \mathbb{N}$ is the next time at which the clock i sounds its alarm. Thus, a state or configuration of the machine is a vector of the form

$$((p_1, t_1), \ldots, (p_n, t_n)).$$

The behaviour of the machine is determined by a transition function, which is a total function that selects sets of instructions to perform on the clocks: we suppose

$$\mathcal{A} : \{0, 1\}^{5n} \to 2^{\{delay(i), lengthen(i): 1 \leq i \leq n\} \cup \{halt\}}.$$

that satisfies $\mathcal{A}(0 \ldots 0) = \emptyset$. Intuitively, this latter condition means that the machine must be asleep until it is woken.

The fact that $\mathcal{A}$'s domain is $\{0, 1\}^{5n}$ means that $\mathcal{A}$'s input is the information of which alarm clocks have alarmed in the last 5 time steps[2] and when they did so. $\mathcal{A}$'s output is simply which clocks to *delay*, which clocks to *lengthen*, and whether the machine *halts* or not. Let $\delta(t)$ denote such a set of actions at time t.

Given a n-alarm clock machine $\mathcal{A}$, and an initial configuration

$$c(0) = ((p_1, t_1), (p_2, t_2), \ldots, (p_n, t_n)),$$

the computation of $\mathcal{A}$ on the given configuration is a sequence

$$c(0), \ldots, c(t-1), c(t) = ((p_1(t), t_1(t)), (p_2(t), t_2(t)), \ldots, (p_n(t), t_n(t))), \ldots$$

of configurations over time t that satisfies for all t,

$$p_i(t+1) = \begin{cases} p_i(t) + 1 & \text{if } lengthen(i) \in \delta(t) \\ p_i(t) & \text{otherwise} \end{cases}$$

$$t_i(t+1) = \begin{cases} t_i(t) + 1 & \text{if } delay(i) \in \delta(t) \text{ or } lengthen(i) \in \delta(t) \\ t_i(t) + p_i(t) & \text{if } t_i(t) = t \text{ and clock } i \text{ alarms} \\ t_i(t) & \text{otherwise} \end{cases}$$

The role of the clocks of the alarm clock machine is to store information on the frequency with which they alarm. In Turing machines the tapes are potentially

---

[2] The number 5 is considered here just because we know the existence of a universal n-clock machine with constant 5. We do not know if there exists a universal n-clock machine exhibiting a smaller structural constant.

infinite, but at any given instant only a finite amount of information is actually stored on the tape. In the same way, the period of the clocks may increase without limit, but at any given instant all alarm clocks have a period bounded by some constant.

Proposition 4.3.2 (Simulation of the *n*-counter machine). *For a n-counter machine that computes in time T, there exists a k-alarm clock machine that simulates it in time $O(T^3)$ with $k \in O(n^2)$.*

In consequence,

Proposition 4.3.3 (Universality of *n*-clock machine). *There is a universal n-clock machine, for some n.*

The observation is that this kind of machine can implement astronomical models of the dynamics of the Solar System in a natural way, partially generalising the way the Stonehenge token game implements the eclipse cycle in a natural way. We can add oracles to the machines, say at precise conjunctions of heavenly bodies. In particular, the model is universal and suitable for thinking about physical bodies, e.g., in the Newtonian gravitational field.

Thesis 4.2. *When we abstract from physical entities, general computers exist in Nature.*

This thesis could be referred to as a *physical principle of general computation.* The standard model of computation, the Turing machine, can be described in an equivalent way, via the n-clock machine model, which resembles the process of making astronomical observations in the manner of the Ancients. Of course, all machine models are abstractions of material components and systems, as are their abstract resources, such as time and space.

The experience of celestial conjunctions described above is not unlike looking at the concept of incomputability as *action at a distance* in the time of Newton. Oracles are needed to fine tune the system once in a while, not only to remove errors (of truncation of real numbers), but also *to remove unpredictability*.

Smith's construction in (2006) can be viewed in the following light: It was known that there were mechanical systems whose asymptotic (long time) behaviour was not known—any computation up to any finite time would hit the problem that just because a given event had not happened yet, did not mean that it would not happen in the future. However, Smith's gravitational machine uses the behaviour of point particles approaching arbitrarily closely to allow uncountably many topologically distinct paths of a point particle in finite time. If you can observe which path the particle actually takes, then you can work out, in finite time, much more information about the initial state of the particles than a numerical error bound. A complication is that there is an infinite set of measure 0 initial states close to the original one where a collision of point particles occurs—a situation where Newton's laws cannot predict the outcome. Smith observed that although we are able to show that Newton's gravitation admits a non-computable orbit, and hence a kind of incomputability of the third kind (3), *special relativity removes it from consideration*.

Expressed in a rather different and radical way, the discovery of a non-computable orbit in Newtonian mechanics refutes Newtonian gravitation theory, *because* it contradicts the physical Church-Turing thesis; in the same way, philosophically speaking, the curvature of light rays from distant stars in the proximity of the Sun refute Newtonian gravitation theory. This adds to the philosophic riddles.

If the reader takes a closer look at the discussion on the Church's Thesis in Odifreddi's text books (Odifreddi 1992, 1999) (e.g., pp. 101–123 in Odifreddi (1992)), he or she will find different formulations of the classical Church's Thesis, such as Kreisel's Thesis M (for *mechanical*) or Kreisel's Thesis P (for *probabilistic*), and so on. Here, at this precise point of his text we find *In the extreme case, any physical process is an analogue calculation of its own behaviour*. And Odifreddi adds a quite interesting footnote:

> *In this case, Church's Thesis amounts to saying that the universe is, or at least can be simulated by, a computer. This is reminiscent of similar attempts to compare Nature to the most sophisticated available machine, like the mechanical clock in the 17th Century, and the heat engine in the 19th Century, and it might soon appear as simplistic.*

In fact, Thesis P states that *any possible behaviour of a discrete physical system (according to present day physical theory) is computable*. Our various systems disprove this (Beggs and Tucker 2006, 2007a,b; Beggs et al. 2007). Also Smith disproves this statement: there exists a Newtonian non-computable orbit. It is not relevant that Relativity Theory removes this pathology: no one would ever believe a few years ago that Thesis P would not be valid. Or is it still valid? Well, physicists say, we don't have point masses or two masses can not come as close as we want. We argue that these are qualitatively physical aspects that are *not in* the formulation of Newtonian gravitation. We will restate saying that no one would ever believe a few years ago that Thesis P would not be valid even for the abstract gravitation theory taught in Physics courses.[3] Is there a student's course notes that considers from scratch the problem of two bodies with non-zero volume. But if then Professor X shows that the n-spherical-body problem gives rise to a non-computable orbit, physicists will say that planets are not really spheres. Here are things in need of better explanations.

Thus, when Cooper and Odifreddi write in (2003):

> *Fortunately, there is another approach—let's call it the "mathematical" approach—which renews the link to Newton. This is a direction rooted in the old debate about whether computability theory has any useful consequences for mathematics other than those whose statements depend on recursion theoretic terminology.*

We would add the direction is also rooted in the *new* debate about whether computability theory has any useful consequences for physics.

---

[3] The most comprehensive study we know is a Treatise about stability of a spacecraft, considering 2-body dynamics, on one side the spacecraft with non-zero dimensions and on the other side the Earth just substituted by its centre of gravity.

We can say that Nature has an algorithmic content: it is greater than the algorithmic content of the Solar System, greater than the algorithmic content of the system Moon–Sun–Earth, greater than the algorithm content of Stonehenge I. Imagine that Stonehenge IV would have been built then, certainly, it would implement the n-clock machine. Does the Universe, or just the universes, have an algorithmic content greater than the algorithm content of Stonehenge IV, abstracting from bounded resources?

We have made a case for the models of computation being intimately related with physical models and physical behaviour. Let us turn to incomputability.

4.4 Incomputability and Predictability: The Discovery of Neptune

Many physical theories provide methods of measurement and calculation. If the calculations are not consistent with measurements then the theory has a problem. The desired measurements are not predicted by the calculations, i.e., they are incomputable. Continuing with the mechanical examples, we will reflect on mechanics, especially Newtonian's theory of gravitation, as a method of calculation. The theme is *preserving and restoring computation* when confronted with incomputabilities.

4.4.1 Action at a Distance

Newton's gravitational law introduced the metaphysical concept of action at a distance. For Newton, action at a distance was done by means of God: space is the *Sensorium Dei* by means of which He stabilizes the system. In his *Opticks*, Newton wrote:

> *…can be the effect of nothing else than the Wisdom and Skill of a powerful ever-living Agent, who being in all Places, is more able by his Will to move the Bodies within his boundless uniform* Sensorium*, and thereby to form and reform the Parts of the Universe, than we are by our Will to move the Parts of our own Bodies. And yet we are not to consider the World as a Body of God, or the several Parts thereof, as the Parts of God. He is an uniform Being, void of Organs, Members or Parts, and they are his Creatures subordinate to him, and subservient to his Will; and he is no more the Soul of them, than the Soul of Man is the Soul of the Species of Things carried through the Organs of Sense into the place of its Sensation, where it perceives them by means of its immediate Presence, without the Intervention of any third thing.*

The removal of *action at a distance* from Physics is not unlike the removal of the Rite in Stonehenge I. The removal of the Rite in Stonehenge I is also like Laplace's removal of the *Sensorium Dei* from Newton's space.[4] But with a *Sensorium Dei* or

---

[4] Newtonian space, like Descartes' substantial space, was not empty but the *Nervous System of God*.

without it, Smith proved the existence of non-computable orbits: an incomputability of the third kind (3), although the proof is based on work of Gerver in (1984), and others for particular cases.

Cooper and Odifreddi raise the question (Cooper and Odifreddi 2003):

> *Why should those without a direct career interest care whether actual incomputability (suitably formalized) occurs in Nature? Even if it did occur, for all practical purposes, how would it be distinguishable from theoretically computable but very "complex" phenomena? Whether chaotic phenomena—such as turbulence—involve complexity or incomputability is interesting, but does it really "matter"?*

The question is also related to Cooper's ideas in his later (Cooper 1999). We think that the answer to this question is not easy.

4.4.2 Waves

Differential equations do exist, having computable coefficients and given computable initial conditions, which cannot be numerically solved by a digital computer: their solution are beyond the Turing limit. Pour-El and Richards provided examples in (1982, 1981, 1979). For example, they considered the three-dimensional wave-equation in Pour-El and Richards (1981). It is well known that the solution $u(x, y, z, t)$ is uniquely determined by the initial conditions u and du/dt at time $t = 0$. They asked whether computable initial data can give rise to non-computable solutions and gave the answer Yes. They gave an example in which the solution $u(x, y, z, t)$ takes a non-computable value at a computable point in space-time.

However, these examples have initial conditions, or boundary conditions, which are not smooth enough to describe real physical situations.

Are all *physical laws* digitally reproducible by a digital computer? If so, then we may talk about non-computable functions as those functions that can not be known through numerical computation using digital computers, despite the fact that they satisfy very simple differential equations. Calculating positions of planets (ignoring some possible incomputabilities suggested in Smith (2006)) was, in fact, a problem of precision. The intrinsically non-computable functions of Pour-El and Richards are of a different kind.

Do we have a model to classify such sources of uncomputability found in Pour-El and Richards (1982, 1981, 1979)? No, we don't. Do you imagine an equation—Poisson's equation—as simple as

$$\psi(x, 0) = f(x),$$

$$\frac{\partial^2 \psi}{\partial x^2} - \frac{\partial^2 \psi}{\partial t^2} = 0,$$

having a non-computable unique solution (non-computable in the sense of conventional computable analysis): there exists not a program such that giving the values of computable numbers x and t with increasing precision will provide $\psi(x, t)$ with increasing precision, despite existing such a program for the function f.

Penrose rejects these examples as useful to a forthcoming *Non-computable Physics*, since the boundary conditions or initial conditions involved are not smooth enough. In Penrose (1989) he stresses this fact before considering the non-computable ultimate physical theory to come and the human mind:

> *Now, where do we stand with regard to computability in classical theory? It is reasonable to guess that, with general relativity, the situation is not significantly different from that of special relativity—over and above the differences in causality and determinism that we have just been presenting. Where the future behaviour of the physical system is determined from initial data, then this future behaviour would seem (by similar reasoning to that we presented in the case of Newtonian theory) also to be computably determined by that data (apart from unhelpful type of non-computability encountered by Pour-El and Richards for the wave equation, as considered above—and which does not occur for smoothly varying data). Indeed, it is hard to see that in any of the physical theories that we have been discussing so far there can be any significant "non-computable" elements. It is certainly to be expected that "chaotic" behaviour can occur in many of these theories, where very slight changes in initial data can give rise to enormous differences in resulting behaviour. But, as we mentioned before, it is hard to see how this type of non-computability—i.e. "unpredictability"—could be of any "use" in a device which tries to "harness" possible non-computable elements in physical laws.*

Computation is preserved by declaring that the boundary conditions are not well-posed physically.

Now, what is the consequence of this to Science? Even for *complex phenomena* like the dynamics of the atmosphere we have strong methods of numerical modelling. We take the Navier-Stokes equation and assume (a) spherical coordinates, (b) that the Earth is not an inertial reference frame, (c) boundary conditions around east North-America's shore and West-Europe's and North-Africa's coast. We presume that (1) such differential equations are integrable by numerical methods and (2) a prediction of the weather for *tomorrow* can be obtained *before tomorrow*. Thus we still have computability considerations and computational complexity considerations. Philosophically speaking, we turn to models of Nature which are predictable. Science in this way is used to make a synthesis of our knowledge about the Universe and to forecast future events. We think that the answers to the questions raised by Cooper and Odifreddi in our last quotation of their article (just before Sect. 4.2) are "Yes", "We don't know", and "Yes". A non-computable world, like the model desired by Penrose, would have a quite different meaning. Assuming that no more computational power is added to computers, we wouldn't have general predictions. The model would be looked upon as divine: suddenly a pattern formation occurs out of the model and some sophisticated computer programs would be able to trace and forecast its trajectory, like a hurricane that although cannot be exactly predicted can be expected and followed, either by satellites or computer programs. A Non-computable Science would be more like a painting in the National Gallery—to look at with respect, admiration, and fascination, being interpreted the critics (would it meet Susan Sontag's *Against Interpretation*). Maybe the questions of Cooper and Odifreddi become:

(a) Does our contemporary science contain patterns of a non-computable model? or
(b) Do we already have a Non-computable Science, hidden in our theoretical achievements? or
(c) Non-computable Science is no more than contemporary fiction, a motor and product of the creative process, like the *stone* was for Alchemy.

No matter the true answer, they make the concluding statement that:

> *Our model says nothing about the mystery of material existence. But it does offer a framework in which a breakdown in reductionism is a commonplace, certainly not inconsistent with the picture given of levels we do have some hope of understanding. It can tell us, in a characteristically schematic way, how "things" come to exist.* [5]

4.4.3 Universes

We also know that *modern science* is losing some coherence and identity. There is just one Newtonian gravitation theory,[6] but with the advent of the General Theory of Relativity, physicists realized that Einstein's beautiful field equation

$$R_{ij} - \frac{1}{2}\, g_{ij}\ R \;=\; \kappa\ T_{ij}$$

could be replaced by different field equations delivering the same realities, delivering the same predicted observations of our Universe. Most probably a non-computable model will deliver also a class of similar observations of the Universe. For instance, is Hoyle's or Hoyle-Narlikar's field of creation *ex nihilo* non-computable? This is not philosophy, since Hoyle's field of creation out of nothing is hard mathematics, although it is refutable nowadays, and not accepted by the scientific community, as the Big Bang Theory is the standard model. Yet, it explains the same observations as the Einstein field equations at some level.

In the 1950s it was a perceived that The Universe was expanding. There was no evidence of the universe showing signs of age. If The Universe was in a steady state then it need to gain hydrogen atoms to preserve the density of hydrogen. Hoyle

---

[5] The reader can have a look at the Eddington's Cosmic Equation, that was a source of explanation of how the Universe came into existence.

[6] Though, in fact, for some time, physicists were tempted to define the law

$$\frac{1}{r^{2.0\cdots025}}.$$

Ridiculous, isn't it? But it worked for a few years, when physicists lost their faith for reasons that will become clear soon.

arrived at the alternative equation

$$R_{ij} - \frac{1}{2}\, g_{ij}\ R\ + C_{ij}\ =\ \kappa\ T_{ij}\,.$$

Associated with the creation tensor C_{ij} was a vector field parallel to a geodesic at each point of the homogeneous and isotropically expanding universe. The field was written

$$C_m = \frac{3c}{a}\,(1, 0, 0, 0),$$

where a is a constant. Hoyle then showed that the solution of the field equations would be given by a metric with space of zero curvature. One can interpret the step as an attempt at preserving the form of Einstein's equations and calculations.

Bondy, Gold, and Hoyle used the word creation rather than formation, just to emphasise the existence of matter where none had been before.

With the Hoyle-Bondi-Gold's model we can evaluate the amount of matter being created at any step of time. But can we predict the point in space where a proton (Hoyle guessed that the spontaneous creation of matter might possibly be in the form of neutrons) will next appear? This is an example of how a non-computable aspect of a theory (we cannot even guess a distribution of matter created[7]) can deliver also computable trajectories of our Universe.

There is a more recent discussion on the evaporation of black holes which is also of relevance here. Hawking (1975) showed that combining quantum field theory with general relativity gave a prediction that black holes should evaporate - that is they radiate particles which results in a loss of mass to the black hole, until eventually the black hole disappears. Similarly to Hoyle's creation field, this particle creation could be viewed as a creation of particles at an essentially random piece of space in the vicinity of the event horizon. This seeming randomness became the subject of a thought experiment: If an observer were to drop an encyclopedia into a black hole, and then wait until it evaporated, would there be enough information in

---

[7] Harrison explains these features in Harrison (2000):

> *There are two kinds of creation: creation of the universe and creation in the universe. On one hand, we have creation (as in cosmogenesis) of the whole universe complete with space and time; on the other, we have creation of things in the space and time of an already existing universe. In the Big Bang universe, everything including space and time is created; in the steady-state universe [of Bondi, Gold, and Hoyle], matter is created in the space and time of a universe already created. Failure to distinguish between the two violates the containment principle. . . The steady-state theory employs creation in the magical sense that at certain place in space at a certain instant in time there is nothing, and at the same place a moment later is something. But the creation of the universe has not this meaning, unless we revert to the old belief that time and space are metaphysical and extend beyond the physical universe; in that case, creation of a universe is in principle the same as the creation of a hazel nut. But in fact uncontained creation (cosmogenesis) is tottally unlike contained creation. Cosmogenesis involves the creation of space and time, and this is what makes it so difficult to understand.*

the particles radiated from the black hole to reconstruct the encyclopedia (Hawking 2005)? This subject of information loss has become hotly debated in recent years.

Cooper and Odifreddi recognizes these different presentations of the Universe stating that:

> *we look for a mathematical structure within which we may informatively interpret the current state of the scientific enterprise. This presentation may be done in different ways, one must assume, but if differing modes of presentation yield results which build a cohesive description of the Universe, then we have an appropriate modeling strategy.*

(See Cooper and Odifreddi 2003) Furthermore:

> *. . . non-locality was first suggested by the well-known Einstein-Podolsky-Rosen thought experiment, and again, has been confirmed by observation. The way in which definability asserts itself in the Turing universe is not known to be computable, which would explain the difficulties in predicting exactly how such a collapse might materialize in practice, and the apparent randomness involved.*

4.4.4 Neptune

The n-clock machine can be implemented with bounded resources in Stonehenge using colored stones, a color for each clock, five colored tokens for each clock.[8] It would have made Stonehenge a huge Observatory (although many existent stones—like the post holes—can handle a large number of calculations that, despite the non-existence of a useful—to the Ancients—implementation of the n-clock machine, make Stonehenge I and II a rather huge Astronomical Observatory). But we are going to talk now about a feature that can not be implemented in Stonehenge: the discovery process!

After Herschell's discovery of *Uranus*, deviations from computed orbit, using Gaussian methods, produced more and more observations of the new slow planet, leading to calculations of more and more accurate orbits. But the new planet always failed to meet the computed orbit: Uranus escaped computation: it was incomputable. If T is Newton's theory of gravity then Uranus was not T-computable.

There are two attitudes. First, one accepts the problem for a period but refuses to give up the calculation. The ancients failed in the prediction of planet cycles. Stonehenge fails as Observatory, but perhaps the memory of the glorious Stonehenge I compels the building of Stonehenge III, the colossal construction of central 3-liths.

Secondly, one accepts that the Newtonian law is wrong and begins the search for the new "true" law of gravitation.

But, as we know, it was too early to reject the Newtonian theory of the Universe. Observations failed; the law $\frac{1}{r^2}$ failed. But Leverrier and Adams, one in Paris, the other in London, proposed that a new planet existed—later called Neptune—to

---

[8] It would be like a Calendar with many entries (cf. Reingold and Dershowitz' *Calendrical Calculations*.

justify departure from predicted orbits and to justify the *true* (Newtonian) law of gravitation.[9] Uranus was again T-computable.

This step cannot be done by a computer program.[10] What is the difference (if the planet was not found) between predicting a planet and replacing Newton's law by another law, being it computable or not? Is the discovery of a new planet a kind of *removal of incomputability*?

How did the scientists respond to the predictions of Leverrier and Adams? They rejected them; they didn't believe them. Is it not a common reaction of an established scientist's mind: *if some hypothesis not in the system is suggested, then it should be immediately rejected*. For example, Airy rejected Adams several times: how, we would like to go to Greenwich and knock at the door to hear him saying *no*! As Morton Grosser tells the story in Grosser (1979), *Airy was an extreme perfectionist, and he divided the people around him into two groups: those who had succeeded and were worthy of cultivation, and those who had not succeded and were beneath consideration [...] Adams solution of the problem of inverse perturbation was thus a direct contradiction of Airy's considered opinion. The Astronomer Royal's negative feelings were indicated by the unusually long time he waited before replying. Airy habitually answered his correspondence by return mail*. In Adams' case he delayed the answer as much as he could.

It would have been enough to look to the sky with a telescope using calculated positions of Neptune.

Feel the pleasure of the following letter of Airy to Adams; it could be adapted to a letter of caution about any thing by an illustrious scientist of our times:

> *We have often trought of the irregularity of* Uranus*, and since the receipt of your letter have looked more carefully to it. It is a puzzling subject, but we give it as my opinion, without hesitation, that it is not yet in such a state as to give the smallest hope of making out the nature of any external action on the planet [...] But [even] if it were certain that there were any extraneous action, we doubt much the possibility of determining the place of a planet which produced it. We are sure it could not be done till the nature of the irregularity was well determined from successive revolutions.*

In a further letter, Airy writes to Adams:

> *We are very much obliged by the paper of results which you left here a few days since, showing the perturbations on the place of* Uranus *produced by a planet with certain assumed elements. The latter numbers are all extremely satisfactory: we are not enough acquainted with Flamsteed's observations about 1690 to say whether they bear such an error, but we think it extremely probable.*
>
> *But we should be very glad to know whether this assumed perturbation will explain the error of the radius vector of* Uranus*. This error is now very considerable.*

---

[9] The difficulties lay with the computation of planet path around the sun without interaction by other planets; Uranus was not computable as a 2-body problem.

[10] Well, Herbert Simon said that it can!—at least Kepler's laws can be rediscovered by computer programs, given Tycho Brahe's data; but not the prediction of a new planet; see Langley et al. (1997).

According to Grosser (1979), on September 18, 1846, Leverrier wrote to Johann Gottfried Galle, assistant to Olaus Roemer. This letter reached Galle on September 23, and he immediately asked his superior, Johann Franz Encke, Director of the Berlin Observatory, for permission to search for the planet. The same night Galle and d'Arrest found the planet: *that star is not on the map*—exclaimed d'Arrest; right ascension $22^h\ 53^m\ 25.84^s$ against the predicted value of Leverrier $22^h\ 46^m$. Although impressive, this accuracy is smaller than Stonehenge's accuracy for the eclipse cycle.

The serendipity of discovery—whether of the kind in the case of Archimedes' *Eureka!*, or in Kepler's laws, or even Kepler's laws according to Herbert Simon's program—are different from the kind of discovery of Neptune.

Thesis 4.3. *We cannot use Natural laws to make a hypercomputer but we can observe objects whose behaviour is hypercomputational.*

The observation of real number values of physical measurements is a starting point for the argument. We don't say that hypercomputation but incomputability is the cause of the discovery of Neptune. Citing Cooper and Odifreddi, *Science since the time of Newton, at least, has been largely based on the identification and mathematical description of algorithmic contents in the Universe. We will look at phenomena—primarily subatomic phenomena—which appear to defy such description*. The "hidden planet" Neptune was a hidden variable that preserves computation.

This incomputability can be seen with the help of the n-clock machine.

We all know that pendulum clocks are quantum systems: each one has exactly two different energy levels, two oscillatory modes: one with the pendulum at rest and other with the pendulum oscillating in a stable orbit. We all know that clocks on the same wall propagate across the wall sound waves, together with their delays or advances, forcing the (coupled) clocks altogether, to a common delay or advance. In Stonehenge, this effect cannot be seen between colored tokens, but on the human machinery that puts the little tokens in motion. Some people forget clocks when they think about the quantum realm. Quantum mechanics in this way also applies to the macroscopic world (of course, not in the sense of making Planck's constant go to zero![11]), in the sense of operators, eigenvectors and eigenvalues.

It works like the ancients, who with the same teleological thinking, finding a disagreement in the predicted Metonic cycle are compelled them to add a further token to the game.

---

[11] It is a good exercise to retrieve

$$\mathbf{F} = m\,\frac{d^2\mathbf{r}}{dt^2}$$

from Schröndinger's equation with $\mathbf{F}$ given by $-grad\ U$, where U is the classical potential field in the original equation.

4.5 Algorithmic Contents of Laws

We disagree with a few statements in Cooper and Odifreddi (2003):

> *In fact ... no discrete model—finite or otherwise—presents a likely host for incomputable phenomena.*

We have at least two exceptions: Wolpert in (1991) studies a discrete neural model with super-Turing capabilities, but with a transfinite number of neurons, and Pollack in (1987) proved that a model of higher-order neural nets is at least universal. Other results on neural networks involve the real numbers. We have an idea that infinite automata can have super-Turing powers, even not involving the real numbers. Secondly, scientifically presenting the Universe with real numbers is not enough to embed in it super-Turing powers. We are always amazed when we hear that a computational model equipped with real numbers allows for hypercomputation. We will start by "defining" super-Turing power of the scientifically presented Universe.

A physical process takes place in time. It is described or understood using specific observable variables which constitute a notion of state. Therefore, the process consists of states that are measurable, either numerically or qualitatively, evolving in time. What makes such a process a computation? The role of initial states and the structure of the observable histories they generate. The role of data in setting initial states and interpreting behaviour as output.

Thesis 4.4. *Up to Turing power, all computations in the Universe are describable by suitable programs, which involve the prescription by finite means of rational number parameters of the system or some computable real numbers; the computations can be generated by a program. Beyond Turing power, we have computations that are not describable by finite means; computations that cannot be generated by any program.*

Computation without a program! When we observe natural phenomena and endow them with computational significance, it is not the algorithm we are observing but the process. Some objects near us may be performing hypercomputation: we observe them, but we will never be able to simulate their behaviour on a computer. What is then the profit to Science of such a theory of computation? The point is that the principle does not tell us about hyper-machines. In this sense hypercomputation can exist. We presume that most of the reactions of the scientific community against hypercomputation are mainly related with *the crazy idea of building a hypercomputer*. We think it is also one of the sources of criticism against the work of Siegelmann (1999) and Siegelmann and Sontag (1994).

But to help the reader to understand that the real numbers alone are not enough to produce any kind of hypercomputation we call upon Analogue Computation.

In the 1940s, two different views of the brain and the computer were equally important. One was the analog technology and theory that had emerged before the war. The other was the digital technology and theory that was to become the main paradigm of computation (see Nyce 1994). The outcome of the contest between these two competing views derived from technological and epistemological arguments. While digital technology was improving dramatically, the technology of analog machines had already reached a significant level of development. In particular, digital technology offered a more effective way to control the precision of calculations. But the epistemological discussion was, at the time, equally relevant. For the supporters of the analog computer, the digital model—which can only process information transformed and coded in binary—wouldn't be suitable to represent certain kinds of continuous variation that help determine brain functions. With analog machines, on the contrary, there would be few or no steps between natural objects and the work and structure of computation (cf. Nyce 1994; Heims 1980). The 1942–1952 Macy Conferences in cybernetics helped to validate digital theory and logic as legitimate ways to think about the brain and the machine (Nyce 1994). In particular, those conferences helped make the McCulloch-Pitts' digital model of the brain (McCulloch and Pitts 1943) a very influential paradigm. The descriptive strength of McCulloch-Pitts model led von Neumann, among others, to seek identities between the brain and specific kinds of electrical circuitry (Heims 1980).

The roots of the theory of Analog Computation lie with Claude Shannon's so-called *General Purpose Analog Computer (GPAC)*.[12] This was defined as a mathematical model of an analog device, the Differential Analyzer, the fundamental principles of which were described by Lord Kelvin in 1876 (see Bowles 1996). The Differential Analyzer was developed at MIT under the supervision of Vannevar Bush and was indeed built in 1931, and rebuilt with important improvements in 1941. The Differential Analyzer input was the rotation of one or more drive shafts and its output was the rotation of one or more output shafts. The main units were gear boxes and mechanical friction wheel integrators, the latter invented by the Italian scientist Tito Gonella in 1825 (Bowles 1996). From the early 1940s, the differential analyzers at Manchester, Philadelphia, Boston, Oslo and Gothenburg, among others, were used to solve problems in engineering, atomic theory, astrophysics, and ballistics, until they were dismantled in the 1950s and 1960s following the advent of electronic analog computers and digital computers (Bowles 1996; Holst 1996). Shannon (in 1941) showed that the GPAC generates the *differentially algebraic functions*, which are unique solutions of polynomial differential equations with arbitrary real coefficients. This set of functions includes simple functions like the exponential and trigonometric functions as well as sums, products, and compositions of these, and solutions of differential equations formed from them. Pour-El, in (1974), and Graça and Costa, in (2003), made this proof rigorous.

The fact is that, although the GPAC model is physically realizable and is an analogue model of some part of the Universe, inputting and outputting real numbers, it does not compute more that the Turing machine, in the sense of Computable Analysis.

---

[12] In spite of being called "general", which distinguish it from special purpose analog computing devices, the GPAC is not a uniform model, in the sense of von Neumann.

Cooper and Odifreddi say that *The association of incomputability with simple chaotic situations is not new. For instance, Georg Kreisel sketched in (1970) a collision problem related to the* 3*-body problem as a possible source of incomputability.*

We think that these ideas are indeed conceived in a few theoretical experiences like in Etesi and Németi (2002) and Xia (1992), although they qualitatively require an unbounded amount of energy,[13] and for this reason, not for theoretical reasons, they are not implementable. Returning to Kreisel, the pure mathematical model of Newtonian gravitation is probably capable of *encoding the halting problem of Turing machines*. This hint is given by Frank Tipler, too, in (1997), based on constructions similar to Xia's 5-body system (in 1992), were we have two parallel binary systems and one further particle oscillating perpendicularly to both orbits. This particle suffers an infinite number of mechanical events in finite time (e.g., moving back and forth with increasing speed). Can we encode a universal Turing machine in the initial conditions? This is an unsolved mathematical problem. Thus, it may well be that the system of Newtonian mechanics together with the inverse square law is capable of non-Turing computations. The hypercomputational power that this system may have is not coded in any real number but *in its own dynamics*. How do we classify such a *Gedankenexperiment*?

In the *Billiard Ball Machine model*, proposed by Fredkin and Toffoli in (1982), any computation is equivalent to the movement of the balls at a constant speed, except when they are reflected by the rigid walls or they collide (preserving global kinetic energy) with other balls, in which case they ricochet according to the standard Newtonian mechanics. The *Billiard Ball Machine* is Universal. Moreover, the faster the balls move, the faster a given computation will be completed. Newtonian physical systems that perform an infinite number of operations in a finite time are well known. Specifically, we just have to consider 4 point particles moving in a straight line under the action of their mutual gravity. Mather and McGee have shown in (1975) that the masses and the initial data of the particles can be adjusted to result in the particles having infinite velocity in finite time. Gerver in (1984) published a paper reporting on a model where, using 5 point particles in the plane moving around a triangle, all particles could be sent to infinity in a finite time.

Can these systems encode hypercomputational sets? We aim at obtaining *either a positive or a negative answer to this question*, i.e., (a) either we will be able to prove that *initial conditions do exist* coding for a universal Turing machine, (b) or we are not able to prove such a lower bound but, we will prove that *encoding of input and output exists, together with adjustable parameters coding for finite control* such that we will have an abstract computer inspired by Newtonian gravitation theory. This result, together with a non-computable character of the n-body problem as shown in Smith (2006) *inter alia*, will turn to be a *strong basis to discuss a possible Church-Turing thesis' violation*. In fact, the non-computable character of the n-body problem is close to Pour-El and Richards' results (Pour-El and Richards 1982), and not so close to a mechanical computer rooted in the structure of the inverse square law.

---

[13] Although the total amount of energy involved does not change.

4.6 Routes to Hypercomputation

Martin Davis published a paper called *The Myth of Hypercomputation* (Davis 2004) which fights against work on hypercomputation in Siegelmann and Sontag (1994); the criticism seems to related with the *dream* of building a hyper-machine. In Siegelmann and Sontag (1994) two paths are started: first, the physical construction of a hyper-machine (that culminates in Siegelmann's controversial claims in *Science* (Siegelmann 1995) that, agreeing with Davis, can be misinterpreted; and, second, the theoretical study of models of hypercomputation, were one searches for neural nets with weights of different types computing diverse computational classes: integer nets are equivalent to finite automata, rational nets are equivalent to Turing machines, polynomial time real number nets are equivalent to polynomial size Boolean circuits, and so on.

In the same way that differential equations in $\mathbb{R}^n$ are used to model Newtonian gravitation, nets with real number weights are worthy of investigation, since for decades engineers have been using them to model learning. In the latter case, philosophical thinking leans towards Davis's considerations. We don't believe in a physical constant L with the value of the halting number.[14] Even if there was, without some reason why it should have that value, we could not use it to make a hypercomputation. Because, if such a constant existed, then we could apply Thesis 4.4. and see objects around us performing hypercomputation having no tool to reproduce it. That would be the case of having hypercomputation as Alchemy, observe Cooper and Odifreddi in (2003):

> *To the average scientist, incomputability in Nature must appear as likely as 'action at a distance' must have seemed before the appearance of Newton's "Principia". One might expect expertise in the theory of incomputability—paralleling that of Alchemy in the seventeenth century—to predispose one to an acceptance of such radical new ideas.*

Alchemy ended and Chemistry started when the *scale* was introduced in Alchemy, a quite good interpretation due to Alexander Koyré. How do we measure hypercomputational behaviour? Suppose we do have a physical constant L having the value of the halting number. Then, if we measure this constant up to, let us say, n digits of precision, for sufficiently large n, and become aware that the program of code e_m halts for input x_m, *how could we verify it?* This would work as a call for observational refutation, it would be, like for Leverrier and Adams, a matter of faith, but in this case without Roemer's telescope in Berlin. Siegelmann's paper in Science looks like Leverrier and Adams trying to convince the scientific community that there is an alien out there. Why was the community not convinced? Well, in first place it seems that nothing in computing escapes mathematical explanation, like Uranus escaped to his computed orbit. But this is not obvious, since sometimes the scientific community do not react as Airy did.

---

[14] Let $\ldots, (e_n, x_n), \ldots$ be an enumeration of programs and natural number inputs. The n-th digit of L is 0 or 1 according to whether or not the program e_n halts on the input x_n.

Do you remember about the scandalous trial in London in 1877? (We learned this from Michio Kaku's *Hyperspace* in (1994)).

A psychic from the USA visited London and bent metal objects at a distance. He was arrested for fraud. Normally, this trial might have gone unnoticed. But eminent physicists came to his defense, claiming that his psychic feats actually proved that he could summon spirits living in the fourth dimension. Many of defenders were Nobel laureates to be. Johann Zollner, from the University of Leipzig came in his defence; so did William Crookes, J J Thompson and Lord Rayleigh. Why this difference of attitude: Airy's reaction to the letters of Leverrier and Adams, with mathematical calculations; Thompson and Crookes reaction to the possibility of psychokinesis working with Zollner?

Newton's *Sensorium Dei* was a metaphysical tool to understand a system of the world that without the intervention of God would collapse in his center of gravity. Leverrier and Adams made people believe again in Newtonian's system of Physics. Departures of computed lunar orbit against observations were explained by Euler. The world is ready for a Laplacian demon to remove God from physical space since Mr. de Laplace *ne besoin pas de cette hypothese* to understand the merry-go-round of the heavenly bodies in the sky. However, what Laplace didn't know is that, most probably, although this system is deterministic, it encodes its own unpredictability and its own incomputability. Probably, not even Laplace's demon has such a computer. In a letter, Cooper observed:

> *… it seems to me that recursion theorists have not until recently really understood or cared what their subject is about, and most still resist even thinking about it (and maybe the same can be said about complexity theorists…). Actually, Gandy was interesting to talk to—as is Martin Davis, of course. We think it is hard for people of my generation and before to adjust to the new fluidity of thinking (or maybe we should say the old fluidity of thinking of the inter-war years).*

The study of hypercomputation should be pursued with mathematics, as with any mathematical concept.

4.7 Final Remarks

Science is a loose federation of different intellectual, experimental and material communities and cultures; the cultures are strong and are not confined to disciplines. We have reflected upon the task of combining our theoretical understanding of computation with that of the physical world. As working scientists, our view is limited to the problems of relating computability and complexity theory with mechanics (Newtonian, relativistic, …). To us there are intriguing questions, observations, theorems and promising approaches.

However, the extraordinary development of the theory of computation since the 1930s has been based on its *mathematical abstractions from the physical world of machines and technologies*. These logical and algebraic abstractions have enabled the rise of digital computation, and have granted Computer Science its intellectual independence from electronics and physics. The mathematical maturity of our abstract theories of computation allow us to look at the physical foundations of computing in new ways. But it also makes our quest more controversial. Beautiful mature abstractions must be traded for clumsy new ill fitting physical notions.

All sorts of questions arise, from a fundamental curiosity about information processing in physical systems, and from a need to understand interfaces between algorithms and physical technologies (e.g., in new problems of quantum information processing, and in old problems of analogue computers). What is the physical basis of computation? Is there a theory of the physical process of making a computation?

But the task of unifying computability and mechanics involves wider issues. One discipline shapes the development of another. To take a example nearby, mathematical logic has had a profound influence of the development of programming languages. Priestley (2008) has examined this process historically and to some extent methodologically using ideas of Pickering (1995). He has shown that there are exciting and rich philosophical phenomena to think about, involving concepts, theorems, practical problems, epistemics and sociology. Combining computability theory and mechanics is a tougher challenge. But it may some essential methodological structures, such as those of "bridging, transcription and filling" suggested by Pickering (1995). Likely it will have new ones, too.

Since the early seventies, we have seen the decline of the enthusiastic debate over what intellectual contribution has computability theory to offer science. The messy debate is back and there new people with a new agenda.

References

Beggs, Edwin J., José Félix Costa, Bruno Loff, and John V. Tucker. 2007. Computational complexity with experiments as oracles. *Proceedings Royal Society Series A* 464: 2777–2801.

Beggs, Edwin J., José Félix Costa, Bruno Loff, and John V. Tucker. 2008a. The complexity of measurement in classical physics. In *Theory and applications of models of computation*. Lecture Notes in Computer Science, vol. 4978, eds. M. Agrawal, D. Du, Z. Duan, and A. Li, 20–30. Berlin/Heidelberg: Springer.

Beggs, Edwin J., José Félix Costa, and John V. Tucker. 2008b. Oracles and advice as measurements. In *Unconventional computing 2008*. Lecture Notes in Computer Science, vol. 5204, eds. C.S. Calude et al. 33–50. Berlin/Heidelberg: Springer.

Beggs, Edwin J., and John V. Tucker. Computations via experiments with kinematic systems. Research Report 4.04, Department of Mathematics, University of Wales Swansea, March 2004 or Technical Report 5-2004, Department of Computer Science, University of Wales Swansea, March 2004.

Beggs, Edwin J., and John V. Tucker. 2006. Embedding infinitely parallel computation in Newtonian kinematics. *Applied Mathematics and Computation* 178: 25–43.

Beggs, Edwin J., and John V. Tucker. 2007a. Can Newtonian systems, bounded in space, time, mass and energy compute all functions? *Theoretical Computer Science* 371: 4–19.

Beggs, Edwin J., and John V. Tucker. 2007b. Experimental computation of real numbers by Newtonian machines. *Proceedings Royal Society Series A* 463: 1541–1561.

Beggs, Edwin J., and John V. Tucker. 2008. Programming experimental procedures for Newtonian kinematic machines. In *Computability in Europe, Athens, 2008*. Lecture Notes in Computer Science, vol. 5028, eds. A. Beckmann, C. Dimitracopoulos, and B. Lowe, 52–66. Berlin: Springer.

Bowles, M.D. 1996. U.S. technological enthusiasm and British technological skepticism in the age of the analog brain. *IEEE Annals of the History of Computing* 18(4): 5–15.

Cooper, S. Barry. 1999. Clockwork or turing universe?—remarks on causal determinism and computability. In *Models and computability*. London Mathematical Society Lecture Notes Series, vol. 259, eds. S.B. Cooper, and J.K. Truss, 63–116. Cambridge, MA: Cambridge University Press.

Cooper, S. Barry, and Piergiorgio Odifreddi. 2003. Incomputability in nature. In *Computability and models*, eds. S. Barry Cooper, and Sergei S. Goncharov, 137–160. New York: Kluwer.

Davis, Martin. 2004. The myth of hypercomputation. In *Alan turing: life and legacy of a great thinker*, ed. Christof Teuscher, 195–212. Berlin/New York: Springer.

Etesi, G., and I. Németi. 2002. Non-turing computations via Malament-Hogarth space-times. *International Journal of Theoretical Physics* 41: 341–370.

Fenstad, Jens E. 1980. *General recursion theory*. Perspectives in Mathematical Logic. Berlin/New York: Springer.

Fenstad, J.E. 2001. Computability theory: structure or algorithms, in W. Sieg, R. Somer, C. Talcott (eds.), *Reflections on the foundations of mathematics: Essays in honour of Solomon Feferman*, Lecture Notes in Logic, volume 15, Association for Symbolic Logic, Poughkeepsie NY, 2002, 188–213.

Fredkin, E., and Toffoli, T,. 1982. Conservative Logic. *International Journal of Theoretical Physics* 21: 219–253.

Gerver, Joseph L. 1984. A possible model for a singularity without collisions in the five body problem. *Journal of Differential Equations* 52(1): 76–90.

Graça, Daniel, and José Félix Costa. 2003. Analog computers and recursive functions over the reals. *Journal of Complexity* 19(5): 644–664.

Grosser, Morton. 1979. *The discovery of Neptune*, New York: Dover.

Harrison, Edward. 2000. *Cosmology, the science of the universe*. Cambridge, MA: Cambridge University Press.

Hartley, Rogers. 1968. *Theory of recursive functions and effective computation*. New York: McGraw-Hill.

Hawking, Stephen W. 1975. Particle creation by black holes. *Communications in Mathematical Physics* 43: 199.

Hawking, Stephen W. 2005. Information loss in black holes. *Physics Reviews D* 72: 084013–084013.4.

Hawkins, Gerald S. 1964. Stonehenge decoded. *Nature* 202: 1258.

Heims, S.J. 1980. *John von Neumann and Norbert Wiener: from mathematics to the technologies of life and death*. Cambridge: MIT.

Hinman, Peter G. 1978. *Recursion-theoretic hierarchies*, Perspectives in Mathematical Logic. Berlin/New York: Springer.

Holst, P.A. 1996. Svein Rosseland and the Oslo analyser. *IEEE Annals of the History of Computing* 18(4): 16–26.

Hoyle, Fred. 1972. *From stonehenge to modern cosmology*. San Francisco: W. H. Freeman and Company.

Hoyle, Fred. 1977. *On stonehenge*. San Francisco: W. H. Freeman and Company.

Killian, Joe, and Hava T. Siegelmann. 1986. The dynamic universality of sigmoidal neural nets. *Information and Computation* 128(1): 48–56.

Kaku, Michio. 1994. *Hyperspace, a scientific odyssey through the 10th dimension*. New York: Oxford University Press.

Kreisel, Georg. 1970. Church's Thesis: a kind of reducibility axiom for constructive mathematics. In *Proceedings of the summer conference at Buffalo N. Y.*, eds. A. Kino, J. Myhill, and R.E. Vesley, 121–150. Amsterdam: North-Holland.

Langley, Pat, Herbert Simon, Gary L. Bradshaw, and Jan M. Zytkow. 1997. *Scientific Discovery, Computational Explorations of the Creative Processes*. MIT Press Ltd, United States, 1987 - location is Cambridge, MA

Mather, J.N., and R. McGehee. 1975. Solutions of the collinear four body problem which become unbounded in finite time. *Dynamical systems theory and applications, Lecture Notes in Physics* 38: 573–597.

McCulloch, W.S., and W. Pitts. 1943. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5: 115–133.

Moore, C. 1991. Generalized shifts: unpredictability and undecidability in dynamical systems. *Nonlinearity* 4: 199–230.

Newham, C.A. 2000. *The astronomical significance of stonehenge*. Coats and Parker Ltd, Warminster (first published in 1972).

Nyce, J.M. 1994. Nature's machine: mimesis, the analog computer and the rhetoric of technology. In *Computing with biological metaphors*, ed. R. Paton, 414–23. London/New York: Chapman & Hall.

Odifreddi, P. 1992. *Classical recursion theory I*. Amsterdam: Elsevier.

Odifreddi, P. 1999. *Classical recursion theory II*. Amsterdam: Elsevier.

Penrose, Roger. 1989. *The Emperor's new mind*. Oxford/New York: Oxford University Press.

Penrose, Roger. 1994. *Shadows of the mind*. Oxford/New York: Oxford University Press.

Penrose, Roger, Abmer Shimony, Nancy Cartwright, and Stephen Hawking. 1997. *The large, the small, and the human mind*, Cambridge/New York: Cambridge University Press.

Pickering, Andrew. 1995. *The Mangle of practice*. Chicago: Chicago University Press.

Pollack, J. 1987. *On connectionist models of natural language processing*. PhD Thesis, Computer Science Department, University of Illinois, Urbana.

Pour-El, Marian B. 1974. Abstract computability and its relations to the general purpose analog computer. *Transactions American Mathematical Society* 199: 1–28.

Pour-El, Marian B., and I. Richards. 1979. A computable ordinary differential equation which possesses no computable solution. *American Mathematical Logic* 17: 61–90

Pour-El, Marian B., and I. Richards. 1981. The wave equation with computable initial data such that its unique solution is not computable. *Advances in Mathematics* 39: 215–239.

Pour-El, Marian B., and I. Richards. 1982. Noncomputability in models of physical phenomena. *International Journal of Theoretical Physics* 21(6/7): 553–555.

Priestley, M. 2008. *Logic and the development of programming languages, 1930–1975*. PhD Thesis, University College London.

Shannon, Claude. 1941. Mathematical theory of the differential analyzer. *Journal of Mathematics and Physics* 20: 337–354.

Siegelmann, Hava T. 1999. *Neural networks and analog computation: beyond the turing limit*. Boston: Birkhäuser.

Siegelmann, Hava T., and Eduardo Sontag. 1994. Analog computation via neural networks. *Theoretical Computer Science* 131(2): 331–360.

Siegelmann, Hava T. 1995. Computation beyond the turing limit. *Science* 268(5210): 545–548.

Smith, Warren D. 2006. Church's thesis meets the N-body problem. *Applied Mathematics and Computation* 178(1): 154–183.

Stoltenberg-Hansen, Viggo, and John V. Tucker. 1995. Effective algebras. In *Handbook of logic in computer science, vol IV: semantic modelling*, eds. S. Abramsky, D. Gabbay, and T. Maibaum, 357–526. Oxford University Press.

Tipler, Frank. 1997. *The physics of immortality: modern cosmology, god and the ressurrection of the dead*. Anchor, New York

Tucker, John V., and Jeffrey I. Zucker. 2000. Computable functions and semicomputable sets on many sorted algebras. In *Handbook of logic for computer science*. Volume V: Logic Programming, eds. S. Abramsky, D. Gabbay, and T. Maibaum, 317–523. Oxford University Press.

Tucker, John V., and Jeffrey I. Zucker. 2004. Abstract versus concrete computation on metric partial algebras. *ACM Transactions on Computational Logic* 5(4): 611–668.

Xia, Jeff. 1992. The existence of noncollision singularities in Newtonian systems. *Annals of Mathematics* 135(3): 411–468.

Wolfram, S. 2002. *A new kind of science*. Champaign: Wolfram Media.

Wolpert, D. 1991. A computationally universal field computer which is purely linear. Technical Report LA-UR-91-2937, Los Alamos National Laboratory, Los Alamos.

Chapter 5
Looking at Water Through Paradigms

A. Perera and F. Sokolić

Abstract Liquid water and its mixtures are used as examples to illustrate the limit of the paradigm of the continuous versus discontinuous description of matter. This is done by going into the details of the modern physical representation of liquids in general and associated liquids in particular. The latter liquids, of which water is the perfect representative, are shown to exhibit local ordering that cannot be described by the current physical theory in a satisfactory way, since these theories are shown to be blind to local order, in the sense that they cannot distinguish ordinary liquids from associated liquids, except by noticing their anomalous properties. The pathway towards a new paradigm is proposed, that bypasses the incoherence of both visions of matter in an unified prospective.

Keywords Paradigm • Physics • Water

5.1 Introduction

How discoveries are made in Science? How does Science formulate and resolve new problems? Such type of questions may be an appropriate starting point from which some interesting epistemological problems can be discerned. This will be illustrated with a specific example from the physical chemistry of liquids, which is our principal field of scientific interest.

A. Perera (✉)
Laboratoire de Physique Théorique de la Matière Condensée (UMR CNRS 7600), Université Pierre et Marie Curie, 4 Place Jussieu, F75252, Paris Cedex 05, France

F. Sokolić
Department of Physics, Faculty of Sciences, University of Split, Nikole Tesle 12, 21000, Split, Croatia

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology, and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_5,

This example can be cast in the terms of the concept of paradigm as developed by Thomas Kuhn in his book “Structure of Scientific Revolutions”. According to Kuhn, a scientist is confronted with two types of problems. The first ones are the standard problems, which Kuhn calls puzzles, and for which the paradigm within which the scientist works is completely satisfactory, which means that a well formed scientist can formulate a problem and find the solution, since he has all the necessary tools for that. The problems of second type are much more difficult, because they need a change of paradigm, in order to be solved. This implies that a new way of formulating the problem and the context in which it is posed, is needed. This may sound vague, but it is difficult to be more specific about it. The most illustrious examples of such change of paradigm are the passage from Aristotelian to Galilean physics, from Ptolemaic to Copernican astronomy, and from classical to quantum mechanics. There are more examples of the change of paradigm, that are less known, but the most important common feature is that a qualitative change of insight is required to solve the problem at hand.

Here, we would like to illustrate the change in paradigms that occurred in the representation of water, which is the most ubiquitous fluid on Earth, not mentioning that this liquid is essential for life and for understanding its origin on Earth. These changes in paradigm are narrowly associated with our conception of matter in the development of human conception of the world, namely in terms of continuous versus discontinuous.

First we introduce water in its complexity, in the sense that it does not quite fit our standard view of a liquid at molecular level. Next, after showing how the statistical representation of matter has historically arisen, we show how one can unify the problematic of water through a presentation that encompasses scales greater than molecular ones, thus merging toward a semi-continuous description of matter. Finally, after recalling how the conception of matter has oscillated throughout the history of ideas, between a continuous and discrete representation, we speculate about the future prospects about this particular liquid that is water.

5.2 The Complexity of Water

We now turn to the problem of water in the liquid state, and more generally to aqueous solutions. Such systems can be studied from two distinct points of view: a macroscopic one (thermodynamics, hydrodynamics), and a microscopic one (quantum chemistry, statistical mechanics). Although water is the most common liquid on earth, and is even quite abundant in glassy ice states in the near vicinity of the solar system, and despite the fact that this particular liquid is very extensively studied (thousands of scientific papers are published per year on this topic alone), there are still several unresolved problems, that mostly concern the second point of view. There are properties of water, which are called anomalous, and which are very difficult to explain in terms of the physics of microscopic phenomena, within a general formalism that is called statistical physics. One such property, for example, is the fact that water has a maximum in density at around the temperature of 4°C. This latter property is narrowly associated to the most commonly known physical fact that ice floats on water, when most solids sink within their corresponding liquids. The main ingredient for these mysterious properties of water is the hydrogen bond. This particular interaction is very different from other interactions between water molecules: it is strongly directional and stronger than the other interactions (van der Waals, electrostatic). As a consequence, water tends to be locally "organized", in the sense that, at any given time, several water molecules are linked by this special interaction. Since it is not the same molecules that are linked to each other as time goes by, a statistical approach is needed. The main question is then: what is the statistical relationship between linked molecules and free ones? It turns out that it is not possible to separate these two states as distinct entities in ordinary liquid water; this has been confirmed experimentally by several methods. Yet, intuitive arguments have led many scientists to think that such separation should exist, in one way or another, since it would help answering many of water's mysterious properties. The next question is then: what exact form these two entities should have in order to both exist and be undetectable by experiments?

It turns out that this very specificity of water is at the origin of the so-called hydrophobic effect, which is of capital importance in biological systems, where water plays an important role. This particular effect describes the fact that water molecules seem to "order" themselves in a specific way around solutes. The exact meaning of "order" is unclear at present, but it would seem that water molecules link to each other by hydrogen bonds differently according to the specific interactions they have with particular parts of the solute molecules-whether these part attract water molecules (hydrophilic) or repel them (hydrophobic). According to this effect, the solutes would experience an effective force that will move them together, or force them to change shape if they can do so, as for example flexible polymers do. Many experimental evidences have led scientists to think that this property of water is at the origin of the particular organization of enzymes and proteins in aqueous solutions. Although a considerable literature has been written on this subject, the exact origin of the hydrophobic effect is still unclear, and contradicting theories have been proposed, that cannot be sorted out on the basis of existing experiments.

Our contribution to this question has come from a general examination of what is called an associated liquid. Many liquids have molecules that can interact through hydrogen bonding, and alcohols are among this category. But their hydrogen bonding is weaker than for water. One of the major recent finding (Soper et al. 2002), was to recognize that mixtures of water and the smallest alcohol, namely methanol, exhibit local immiscibility, while being macroscopically fully miscible. This was suspected for a long time from experimental results, but the recent confirmation from computer simulations has opened new horizons. Why? In short, local immiscibility should, in principle, lead to an inevitable global instability, and thus break the homogeneity of the mixture and lead it to global demixing. It turns out that it is not easy to explain by using current theoretical material how a macroscopically homogeneous system can sustain micro-immiscibility. Let us examine in some details why this happens.

5.3 A Statistical View Point on Water

The macroscopic behaviour of matter, and liquids in particular, is governed by mechanics and thermodynamics. This latter science was fully elaborated one and a half centuries ago without the need to refer to the molecular nature of matter. It even incorporated mathematical relations to express the global stability of matter-in a mechanical sense. It is the very existence of such relations that opened the route to look beyond the continuum of matter, into the deeper and finer structure of it. Maxwell, and more importantly Boltzmann made an outstanding breakthrough by discovering how simple hypothesis on the ballistic nature of thermally agitated molecules could lead to the very fundamentals of the thermodynamical laws. This was a change in paradigm analogous to the second kind mentioned above. The probabilistic approach introduced a new insight, but also a new calculational approach to describe matter. One of the major concepts that was introduced, is that of the "correlation function", which describes how molecules are ordered next to each other, *in average*. In the present version of the statistical description of liquids, the hierarchy of required knowledge is as follows: one prescribes the molecular interactions, which in turn allow to calculate the correlation functions, the latter which permit to compute thermodynamical properties of macroscopic matter. When correlations become long ranged, matter becomes less stable in its present state -or phase, and a transition occurs to a new phase. The growth of long range correlations is intuitively understood as follows: molecules that are far apart become more and more "aware" of each other by an *indirect* process. This could be considered as an *effective* interaction between those far apart molecules. This description applies perfectly to many kinds of phase transitions and it is the proof that this description is robust.

Now what happens when mixing associated and inert liquids? The hydrogen bonding directional interaction "ties" several of the associated molecules into a cluster and the inert molecules are grouped in remaining interstitial spaces: this is called microheterogeneity, or microsegregation of species. There are several ways how associating molecules can bind to each other, and this depends on the geometry of their interactions and constituent atoms: this is what distinguishes water from other associated liquids, for example. What happens now to the correlation functions? these functions see now a micro-domain of self-clustered molecules as being strongly correlated, but not beyond a certain range which corresponds to the size of the microheterogeneity. This is what distinguishes this type of correlations from those related to stability, where these correlations are truly long ranged. The distinction between these two types is very important, even if they may look similar within the first few molecular shells. All this is most conveniently formulated mathematically in the reciprocal Fourier space and by the detection of specific peaks in the dual function to the correlation function, called structure factor. It is the interplay between concentration fluctuations, related to the stability of the granular matter, and the clusters, related to specific type of interactions, that creates the delicate balance between micro-immiscibility and macroscopic miscibility.

5.4 A Different View on the Organisation of Matter

Our contribution here has been to separate the two phenomena in the mental representation of matter through correlation functions. We have introduced a new step in the representation of matter, namely that of microheterogeneity as an intermediate state between the microscopic interaction and macroscopic phase (Perera et al. 2007; Zoranic et al. 2007). We would like to see how the actual formalism of statistical mechanics allows one to characterize this state when specific interactions are prescribed, such as hydrogen bonds.

This view point is somewhat in variance with the actual credo of phase transitions, where the details of the interactions do not matter when dealing with macroscopic phenomena that drive the system out of stability, namely phase transitions. In this theory, called Renormalisation Group Theory, density or concentration fluctuations play an essential role in driving phase transitions. When such fluctuations are about to destabilize the current phase and create a new phase, the details of the interactions do not matter any more. In our approach, however, specific interactions can block the spread of fluctuations by condensing locally into specific shapes. Indeed, it is well known that hydrogen bonding increases the stability of the liquid state up to very high temperatures. For example, the critical temperature of water is about 640K, but if we remove the hydrogen bond interaction this critical temperature falls down below 200K. The importance of specific interactions in destroying phase transitions is not well acknowledged. There exists however a model where this effect is fully demonstrated: it is the dipolar hard sphere model. In this model, there is a strong directional interaction that comes from the fact that dipoles preferentially lie head-to-tail, forming long chains. On the basis of *average* interaction, one would predict a gas–liquid phase transition, if fluctuations alone were driving the stability of this system. However, it turns out that the local formation of chains suppresses this transition, or possibly drives it to a much lower temperature than would be predicted on fluctuation analysis alone. The chains here play the same role of the microheterogeneity in water and aqueous system. It is very important to note that these chains have no definitive existence in time, these are essentially fluctuating entities, evanescent in time. Yet, the existence of these "organised fluctuations" affects dramatically the stability of this fluid. This example illustrates very well the deep interplay between fluctuations and microheterogeneity.

It turns out that this approach is a conceptual leap that has interesting resonances in other domains of physics as well, namely those that concern the appearance of local patterns in the midst of a homogeneous *ordered* state, namely low temperature and high energy physics. Our current task is to understand the difference between the role of the ordering field in creating local patterns, as opposed to the absence of such order in our case, yet with similar local objects. The key to a possible unification of the two branches of physics might be the idea of local *symmetry invariance*, not in the global state, but in the local level of representation.

5.5 Conclusion: By-passing the Traditional Dichotomy of Continuous Versus Discontinuous Representation of Matter

Historically, our representation of matter has oscillated between a continuous and a discontinuous description. These two descriptions are associated with the macroscopic versus microscopic view point over matter. There has been a permanent struggle between the continuist and the atomistic world view during the whole history of physics and chemistry (natural philosophy). Although matter appears continuous to our senses, the idea that matter is ultimately composed of indivisible components, called atoms, was proposed already by philosophers of the Ancient Greece in the fifth century BC. This idea was considered to be seducing because it allowed a simplification of the description of the world. For example, for a continuist, liquid water and ice are very hard to rationalize, while for an atomist, they are only different state of the same molecular matter. Conversely, the logical chain of the search for more fundamental constituent is brutally stopped in the atomistic view, without possibility to investigate further below, thus appearing somewhat superficial representation of the world. So, this whole struggle can be resumed in the following way: atomists propose a asymmetric view of the world, since it can be explained from above but not from below, while continuists propose a overly complexified view of the world without unifying principles.

Let us now illustrate this in the case of water and its mixtures with other liquids. Thermodynamics is the perfect example of a continuous description of matter. In the end of nineteenth century, it was even considered as the perfect paradigm for the description of matter. Thermodynamics explains perfectly the state change between liquid and solid, without the need of any molecular picture. This was even considered as a standard to illustrate the inner beauty of this formalism (*Pierre Duhem and Wilhelm Ostwald*). However, thermodynamics cannot explain the phase change between a monomer solution and a micellized solution, although it can detect that a phase change has occurred. This is because the two phases have the same symmetry: they are both globally disordered. In short, thermodynamic is blind to symmetrical changes. This type of phase change can only be rationalized if monomers and interactions are explicitly introduced, which is then a molecular representation of matter. There are many such other examples. Phase changes with symmetry have been beautifully formalized by the Russian physicist and

Nobel prize winner Lev Landau. This formalism is an atomistic view point, and explains beautifully the above mentioned phase transition between monomer and micellar solutions. It is now the new dominating paradigm, encompassing classical thermodynamics. However, this new formalism cannot distinguish between an homogeneous liquid and a micro-heterogeneous liquid, because this difference cannot be explained in terms of *global* symmetry changes, since this symmetry is the same in both cases, but only in terms of *local* symmetry. So, this new paradigm is as blind in front of these two states differing by local symmetry, as thermodynamics was in front of the two phases mentioned above.

We think that Landau atomistic level paradigm should be, in turn, over-passed by a formalism that is not blind to the difference in local order. In order to see the local order, we need to get above the molecular picture in order to see the microheterogeneity of organized matter. This view point, while still being atomistic, encompasses now both visions. We consider that it by-passes the incoherence of both visions, on the basis of an existing reality and the need to describe it.

References

Perera, A., F. Sokolic, and L. Zoranic. 2007. Micro-structure of neat alcohols. *Physical Review E* 75: p060502R

Soper et al. 2002. Molecular segregation observed in a concentrated alcohol-water solution. *Nature (London)* 416: 829

Zoranic, L., F. Sokolic, and A. Perera. 2007. Microstructure of neat-alcohols: a molecular dynamics study. *The Journal of Chemical Physics* 127: 024502

Chapter 6
Introducing Universal Symbiogenesis

Nathalie Gontier

Abstract One of Neurath's ambitions was to increase the uniformity of scientific languages. A modern day attempt to obtain this goal of a uniform scientific language can be found in the field of evolutionary epistemology. Evolutionary epistemologists are characterized by their quest for universal formulas of evolution that can explain evolutionary change in a variety of phenomena. The most known are universal selectionist accounts. The latter are introduced and implemented within philosophy of science and extra-philosophical fields alike. But what about other evolutionary theories such as symbiogenesis? The process of symbiogenesis need not be confined to either the microcosm or the origin of eukaryotic beings. On the contrary, just as natural selection today is being universalized by evolutionary biologists and evolutionary epistemologists, so symbiogenesis can be universalized as well. It will be argued that in its universalized form, symbiogenesis can provide: (1) a general tool to examine various forms of interaction between different biological organisms, and (2) new metaphors for extra-biological fields such as cosmology, the cultural sciences, and language. Furthermore universal symbiogenesis can complement, if not provide an alternative, for universal selectionist accounts of evolution. As such, universal symbiogenesis can provide a scientific language that enables more uniformity between different disciplines.

Keywords Evolutionary epistemology • Universal selectionism • Universal symbiogenesis

---

N. Gontier (✉)
Centre for Philosophy of Science, Universidade de Lisboa, Lisbon, Portugal

Marie Curie Actions, Centre for Logic and Philosophy of Science,
Vrije Universiteit Brussel, Brussels, Belgium
e-mail: Nathalie.Gontier@vub.ac.be

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology,
and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_6,

6.1 Introduction

In his "An international encyclopaedia of the unified sciences", Otto Neurath (1936) contemplated on how different sciences use different scientific languages. The cacophony of scientific languages makes it difficult for scholars working within these different sciences to communicate adequately with one another and to surpass their disciplinary boundaries. Neurath argued that we should strive to increase the uniformity of scientific languages.

Classically, languages that enable such increase in uniformity are associated with formal languages, more specifically the language of mathematics or logic. More recently, the language of evolution also takes on such unifying proportions. Since the rise of evolutionary epistemology (Campbell 1974) and evolutionary psychology (see Barrett et al. 2002 for an overview), a variety of phenomena are understood from within a selectionist framework. The theory of natural selection is applied to the evolution of scientific theories (Campbell 1997; Hull 1988; Plotkin 1995; Toulmin 1972), culture (Blackmore 1999), language (Pinker and Bloom 1990) and the brain (Changeaux 1985; Edelman 1987).

"Universal" selectionist/Dawinian accounts (Cziko 1995; Dawkins 1983) have turned out to be very fruitful in uniting, at the very least, the humanities or life sciences with the biological sciences. The great success of this selectionist approach is partly due to the fact that a "universal" jargon is provided to researchers working within different disciplines and faculties. The fact that numerous disciplines flourish by the adoption of the universal selectionist approach makes one wonder whether other evolutionary theories can undergo the same faith. Can other evolutionary theories also be universalized and serve to study not only biological processes but also extra-biological processes? In other words, can other evolutionary theories also provide the lexicon for a universal, unifying scientific language?

Here it will be argued that the latter is indeed the case and one nonselectionist evolutionary theory, symbiogenesis, will be universalized.

6.2 What Is Evolutionary Epistemology (EE)?

Evolutionary Epistemology is a discipline that investigates how evolutionary theory can be applied to study the evolution of phenomena such as cognition, knowledge, science, culture or language. Beneath, a short summary of the basic tenets of EE are given, followed by a distinction between older and younger branches of the field.

6.2.1 The Traditional Goal of EE

Originally, EE focussed on how evolutionary theory, especially natural selection theory, can inform scholars on how we and other organisms gain knowledge. According to the Stanford encyclopedia of philosophy,

> Evolutionary Epistemology is a naturalistic approach to epistemology, which emphasizes the importance of natural selection in two primary roles. In the first role, selection is the generator and maintainer of the reliability of our senses and cognitive mechanisms, as well as the "fit" between those mechanisms and the world. In the second role, trial and error learning and the evolution of scientific theories are construed as selection processes. (Bradie and Harms 2004: 1)

Firstly, natural selection can be understood not only as the maintainer but also as the generator of the reliability of our senses and cognitive mechanisms. This is because adaptation results in a fit between the senses and cognitive mechanisms and the outer world. Secondly, natural selection can also serve as a metaphor to explain trial and error learning and the evolution of scientific theories.

The first idea is called the EEM program (evolution of epistemological mechanisms), while the second idea corresponds to the EET (evolutionary epistemology of theories) program of evolutionary epistemology (Bradie 1986). Traditionally the EEM program is understood to be normative. Because natural selection is the generator of our cognitive mechanisms, cognition can only be explained by making use of natural selection. The EET program on the other hand is understood to be merely descriptive. It endorses the view that the evolution of life by means of natural selection and the evolution of science undergo analogous processes. Although no causal forces of natural selection are a priori assumed to underlie both the evolution of science and life, this is not a priori excluded to be the case either.

Early evolutionary epistemologists particularly focused on traditional philosophical questions concerning knowledge. Problems investigated include the reliability of our senses, the reference problem, and synthetic a priori claims (see the works of Lorenz 1941; Popper 1963; Toulmin 1972; Munz 2001). What distinguishes an EE-approach to knowledge acquisition from classic epistemology is elucidated by Munz (2001: 9). Scholars who worked within a "first philosophy", assumed that knowledge concerns the relation between a human knower and something known (according to rationalists) or knowable (according to empiricists). With the emergence of a sociology of knowledge, it is the relation between different human knowers that is under investigation. The possibility of there existing a knowledge relation between a knowing agent and the outer world becomes questionable. Rather, it is argued that regimes define what counts as knowledge and what does not. Within EE, however, knowledge becomes defined as the relation between all biological organisms and their environment. Natural selection justifies such a relation, because organisms that survive are adapted to the environment. Fish for example, can be argued to provide knowledge of, or even a theory of water, because fish are adapted to an aquatic environment. Fitness can therefore become a measure for correspondence between an organism and the environment. In this regard, both physical traits as well as the products of physical traits, such as cognition, culture, or language, need to be examined from within evolutionary theory.

6.2.2 Is This the Only Role EE Can Play?

EE has mainly been occupied with the "evolutionizing of epistemology"; natural selection theory is used to analyze knowledge. But EE can also fulfil another role which involves the "epistemologizing of biology" (Callebaut and Pinxten 1987: 17). What does this mean?

One of the founding fathers of EE, Donald T. Campbell (1974) indicated that natural selection not only works within the evolution of life, it is also active within the evolution of cognition. He abstracted a template of natural selection which he called blind-variation-and-selective-retention. According to Campbell, this template or scheme can serve as a heuristic that informs scholars on how natural selection can be active in the evolution of life, cognition, or language.

By introducing his blind variation and selective retention scheme Campbell was implicitly raising the following questions: how do you study an evolving entity properly, what is the right methodology?

The question of how to study evolution methodologically was also asked by scholars working within the units and levels of selection debate. The concept "unit of selection" was first introduced by Lewontin (1970). At that time, this concept did not relate to the then prevailing EE debate that Campbell had introduced. On the contrary, Lewontin wrote his article to tackle Hamilton's idea of the existence of group selection (for a more elaborate overview, see Gontier 2006c). But interesting is that, although unfamiliar with Campbell's blind variation and selective retention scheme, Lewontin (1970, 1) too abstracted what he calls *"a logical skeleton"* of natural selection. This skeleton, "*phenotypic variation, differential fitness and heritability of that fitness"* can be applied to *"different units of Mendelian, cytoplasmic, or cultural inheritance"*. This skeleton can thus be universalized and it can explain the evolution of certain biological and cultural processes. Moreover, Lewontin argued that this skeleton also sets clear boundaries on what can be understood as a unit of natural selection. Something can only evolve by means of natural selection if the conditions set out in the skeleton are met. Lewontin therefore also methodized biology: he introduced an epistemic framework that allows one to investigate what the units of selection are, and how evolution by means of natural selection occurs.

Lewontin's paper was highly influential, and lead to a series of methodological questions about the application range of natural selection and how the mechanism proceeds.

In 1976, Dawkins wrote his book *The Selfish gene*, and in it he argues, contrary to the Modern Synthesis, that the unit of selection is not the phenotype, which he regarded as a mere vehicle,[1] but the gene. Genes, as units of selection, can be universalized into replicators. In his 1983 paper, entitled *universal Darwinism*, Dawkins claimed that anywhere in the universe where evolution occurs, this evolution would entail the selection of the replicator. A replicator is defined as "any entity in the universe of which copies are made" (Dawkins 1982: 162). He further raised the possibility that there also exist cultural replicators, memes, that evolve by means of natural selection.

Dawkins as well as Lewontin were formulating their ideas exclusively within evolutionary biology and were at that time unaware of the discussion going on within EE. Although Campbell already formulated his template of natural selection in the late 1950s, these ideas were not known within evolutionary biology. These endeavours would become combined by Brandon (1982) who wrote an article entitled *The levels of selection*, and Brandon and Burian (1984), who together published an anthology that combined all the different endeavours that set forth to find universal methodologies to study life at all its ranks, as well as the products of life such as cognition, language and culture.

From then onwards, both evolutionary epistemologists and scholars working within the units and levels of selection debate would engage in finding systematic ways to identify the various units and levels of selection, as well as how the mechanism of natural selection occurs. This search for a universal methodology to study evolution by means of natural selection at all levels of reality is best characterized as the "epistemologizing of biology".

6.2.3 Current Trends

If we investigate current trends, three observations can be made.

Firstly, the evolutionizing of the field of epistemology, as set out by both the EET and EEM program, has primarily been conducted from within a selectionist, adaptationist framework. It is investigated how natural selection shapes our cognitive devices, and how adaptive the latter are.

Within current evolutionary biology however, the adaptationist approach has been either complemented or criticised by numerous different evolutionary theories. The most important ones are neutral theory, (developmental) systems theory, symbiogenesis and (see Schwartz 1999 for a good overview).

Only some of these theories are finding their way into evolutionary epistemology. Wuketits (1990, 2006) has introduced a non-adaptationist approach to cognition, Riedl (1984) has developed a contructivist approach to knowledge, and system

---

[1] David Hull (1980, 1981, 1988) would later on again counter this idea by introducing the universal notion of an *"interactor"* that was to be understood as the proper unit of selection. The interactor largely corresponds with the phenotype that is traditionally understood to be the unit of selection and it is selected at the level of the environment.

theoretical approaches have been introduced by Hahlweg and Hooker (1989). These latter theories can be characterized as a form of "new EE" that can be distinguished from traditional, selection-focussed EE (Gontier 2006a).

Central to these new forms of EE is that they focus on the active role that an organism can play during its evolution. Although an oxygen-breathing human, for example, might provide a theory on the atmosphere that surrounds it, this human being does not provide an adequate theory about the outside temperature. This is because humans are, to a certain extent, able to actively construe a niche for themselves. Certain internal systems, such as hormones, regulate body temperature, and humans can also wear clothing to protect them from the external temperature. These traits do not allow one to assume that there is a complete 1:1 correspondence between humans and certain aspects of their environment. Organisms, more often than not, are not fully adapted to their environment (Gould and Lewontin 1979; Lewontin 2000) and therefore only provide partial knowledge and theories about their environment. New EE tries to incorporate these findings in theories on knowledge. But although constructivist and system theoretical views have been implemented, various evolutionary mechanisms currently remain understudied.

Secondly, selectionist EE goes beyond the study of philosophical issues of knowledge. Traditionally, the application of evolutionary theory outside the biological realm into areas such as philosophy, anthropology, psychology, linguistics or economics was understood to merely be of a descriptive nature. Now however, EE is regarded as a normative discipline that enables the study of philosophical and extra-philosophical phenomena alike by making use of evolutionary theory. A shift has taken place where epistemology went from being the subject of study to being the provider of a new, on evolution-based methodology.

Because of their engagement in the units and levels of selection debate, evolutionary epistemologists study how on the one hand, genes, organisms, groups or species, and on the other, culture, language, or social norms evolved by natural selection. The application of selectionist accounts to domains such as language and culture raises the question of how many units and levels of selection there really are and how exactly natural selection works on these units and at what levels. In light of this, numerous selectionist formulas have been developed. There is the blind-variation-and selective-retention scheme, first introduced by Donald Campbell (1974, 1997); the generate-test-regenerate scheme developed by Plotkin (1995), and the replication-variation-environmental interaction scheme, advanced by Hull (1980, 1981, 1988). For a discussion of all these different formulas see Cziko (1995) or Gontier (2006c).

This "epistemologizing" of natural selection has introduced a new framework, called "Universal selectionism" (Cziko 1995), that claims that all and only selectionist accounts are valid when studying the evolution of both life and non-life. As such, EE functions as a unifying language. Although, to my knowledge, nobody is drawing the relation with the unity of science debate and Neurath's plea to make different scientific theories more uniform, universal selectionist formulas today are taking on the role of such a universal, unifying scientific language. Universal selectionist formulas are offering a uniform methodology that can be equally applied by biologists as well as sociologists, anthropologists or linguists.

Thirdly, because of the success of selectionist approaches, the "epistemologizing" of theories other than natural selection has by and large been neglected, and some adherents of universal selectionism go so far as to claim that such epistemologizing is unnecessary. Yet, if we take evolution seriously, we also need to investigate how these theories possibly provide insight in both biological as well as sociocultural evolution.

Here, it is defended that scholars should engage in epistemologizing all known evolutionary mechanisms. We need to abstract a logical skeleton of all known evolutionary mechanisms and identify the units and levels whereupon these mechanisms work. This allows for a systematic way to study how evolutionary mechanisms underlie both the evolution of life and products of life such as language, cognition or culture. Abstracted skeletons can serve as a template or heuristic, that informs scholars working in both the life and human sciences on how a mechanism proceeds, and where, at what ranks of reality it occurs.

These future templates will also serve as a universal language that enables better communication between various disciplines such as evolutionary biology, evolutionary linguistics, evolutionary psychology and evolutionary anthropology. Skeletons can serve as universal templates that test the possibility that certain phenomena evolve by means of a certain mechanism, and they can serve as templates that allow scholars to model the evolution of those phenomena. In other words, these templates or skeletons can serve as a universal language that enhances interdisciplinarity between the various sciences in exactly the way Neurath intended it.

6.3 Epistemologizing Symbiogenesis

Given the merit of the selectionist approach, it is legitimate to ask whether other evolutionary theories, such as symbiogenesis, can also provide universal templates that can serve as methodologies to study aspects of both biological and sociocultural evolution. In the remainder of this article, we first examine the basic principles of symbiogenesis. Secondly, we abstract a universal template of symbiogenesis, and thirdly, we implement this template in both biological as well as sociocultural evolution.

6.3.1 Symbiogenesis

Symbiogenesis is an evolutionary mechanism that explains the origin of eukaryotic beings. Its modern version is formulated by Lynn Margulis (1999), Margulis and Sagan (2000), Margulis and Dolan (2002), but the ideas on symbiosis and symbiogenesis date back to the early 1900s and were first formulated by Constantin Mehrezskhovsky (for an overview see Sapp 2003; Sapp et al. 2002).

All living organisms are classifiable according to two cell types, prokaryotes and eukaryotes. All unicellular organisms are prokaryotes. Characteristic of the cells of these organisms is that they do not contain a nucleus. Eukaryotic organisms, on the other hand, are all organisms whose cells contain a nucleus that in turn contains the chromosomes that package the genes. Only the first kingdom of life, the bacteria and Archaea, are made up of prokaryotic beings. All other kingdoms, i.e. the protists, fungi, animals and plants, are eukaryotic beings.

Besides a nucleus, an eukaryotic cell often contains organelles. Organelles are little cell bodies that contain their own genetic material. These organelles (e.g. mitochondria and plastids) are only inherited from the mother.

The serial endosymbiotic theory explains both the origin of eukaryotic beings as well as the organelles that reside inside them. The theory demonstrates how once free-living prokaryotes entered each others bodies. Through these mergings they evolved into nucleated, eukaryotic beings. And also the organelles of these eukaryotic organisms are descendents of once free-living prokaryotes that merged with early life forms.

More specifically, two billion years ago, spirochetes and thermoplasma merged. Somehow, spirochetes must have penetrated thermoplasma and they must have become trapped in these new hosts who in turn were unable to digest or delete this newly acquired material. This merger resulted in the emergence of nucleated organisms called protists. Thus symbiosis (the living in close contact or merging of different species into one another) lead to symbiogenesis (the emergence of new biological entities through permanent symbiosis). Some of these protists further merged with bacteria known as paracocci. The paracocci that entered these early protists evolved into organelles called mitochondria. The cells of most multicellular life contain mitochondria. They allow a cell to breath. Plant cells also contain chloroplasts, i.e. organelles that are involved in photosynthesis and that are responsible for the green colour that plants have. These chloroplasts, SET-theory shows, are the descendents of once free-living cyanobacteria.

Thus, bacteria, or bacteria and eukaryotic organisms can merge and these horizontal mergings can result in the evolution of new structures if not new species. The latter process is called symbiogenesis. Contrary to the Modern Synthesis that emphasizes that evolution occurs vertically, through speciation, SET theory demonstrates that evolution can occur through the horizontal mergings of different lineages.

6.3.2 Universal Symbiogenesis

The physicist Freeman Dyson (1988, 1998, 1999) was the first to universalize symbiogenesis. He used symbiogenesis to explain how life originated on this planet. His theory, known as the *"double origin theory"*, synthesizes two competing views on how life evolved on earth. Spontaneous generationists such as Oparin (1955) or Fox and Dose (1972) argue that life evolved from spontaneously evolved protein-like structures and/or cells. RNA only evolved later, within these protein-like structures. Adherents of an RNA world (e.g. Eigen and Schuster 1977; Eigen 1996; Gilbert 1986; Orgel 1994), on the contrary, assert that genes evolved first. These two competing views are united by Dyson (1998) through symbiogenesis. He argues that Eigen's autocatalytic molecule could only develop inside protein-like cells as the first parasite of this protein-like life. Life thus did not originate once but twice: once as metabolism and once as information. And metabolism and information, respectively protein cells and genes, got combined symbiotically. Dyson (1988: 81):

> I am suggesting that the Oparin and Eigen theories make more sense if they are put together and interpreted as the two halves of a double-origin theory. In this case, Oparin and Eigen may both be right. Oparin is describing the first origin of life and Eigen the second.

By applying the mechanism of symbiogenesis to the molecular level, he brings symbiogenesis into the physical and chemical realm and as such he broadens the explanatory scope of symbiogenesis. But Dyson (1998) goes further than that. Besides its role in the evolution of life, he uses an abstract principle of symbiosis to explain both processes in the universe (such as galactic cannibalism, symbiotic stars, black wholes) as well as processes in science (e.g. the synthesizing of theories).

According to Dyson, two evolutionary phenomena occur universally: speciation events and symbiosis events. Both lie at the basis of order-disorder transitions, differentiation, and the emergence of new structures. And all occur during rapid phases in history which are punctuated by long periods of stasis. Dyson thus does for symbiogenesis what evolutionary epistemologists have done for natural selection. That is, he universalizes the principle so that it can be used to explain a variety of different phenomena. As such, he provides different scholars with a universal language. Dyson (1998: 121) defines universal symbiogenesis as:

> . . . The reattachment of two structures, after they have been detached from each other and have evolved along separate paths for a long time, so as to form a combined structure with behaviour not seen in the separate components. (Dyson 1998: 121)

This definition is very useful to explain certain physical and biological phenomena. However, this definition is also somewhat biased towards physical and biological phenomena. There is no need whatsoever to a-forehand assume that symbiogenesis can only occur when structures have first become detached before merging, or that only two structures can engage in symbiogenesis. Especially cultural ideas, scientific movements or languages are often made up of a variety of entities that somehow got combined. In other words, if we adjust Dyson's definition of universal symbiogenesis, we might be able to also explain extra-biological phenomena such as language and cultural evolution. We might thus be able to universalize the language even more and provide a universal methodology in Neurath's sense.

What is basic to symbiogenesis? And what should a logical skeleton of symbiogenesis consist of (the following is paraphrased from Gontier 2007: 174–175)?

(1) Basic to symbiogenesis is *interaction*. We are looking for a universal formula of symbiogenesis that is applicable to as many different phenomena as possible. Therefore it is wise not to specify a-forehand the type of interaction (mutualism, parasitism, symbiosis, commensalism), the type of entities that interact (individuals, lineages, traits) or the number of entities that interact. Specification would lead to exclusion and the goal here is to find as many possible types of symbiogenesis as possible.
(2) Symbiogenesis also implies *horizontal mergings* that lead to *permanent* and *irreversible* changes, which form the basis of evolutionary *novelty*. Carrapiço and Rodrigues (2005: 59060R-2) reason that symbiogenesis is an evolutionary mechanism that unfolds through symbiosis: *"… symbiogenesis should be understood as an evolutive mechanism and symbiosis as the vehicle, through which that mechanism unfolds."* It is important to emphasize that once something emerged by symbiogenesis, it can (further) evolve vertically. However, the latter is always preceded by a horizontal merging of different entities. The merging is permanent and irreversible because the merged entities become one new entity that demonstrates behaviour not seen in the separate components.
(3) Such horizontal mergings can occur *rapidly*, and are always *discontinuous*. They are discontinuous because the merging results in the evolution of something new.
(4) Finally, as Margulis (1999) emphasizes, symbiogenesis entails *"individuality through incorporation"*. A new distinctive entity emerges *only* through the interaction of other entities.

This logical skeleton can be universalized into the following formula of universal symbiogenesis (Gontier 2007: 174–175):

> Universal symbiogenesis occurs when new entities irreversibly and discontinuously evolve out of the horizontal merging of previously independently evolving entities.

Because the goal is to delineate different phenomena as symbiogenetic, including phenomena that were previously not identified as such, the above definition is kept unspecified. Nonetheless, we can call the interacting agencies or entities *symbionts*. In biology, examples of symbionts are viruses, bacteria or parasites that interact with their host. When such horizontal interaction leads to the emergence of a new entity, this entity is called the *symbiome* (Sapp 2004: 1047). The symbiome can again become a possible unit of horizontal or vertical evolution (including evolution by means of natural selection).

We can also provide an evolutionary epistemological, universal dimension to the concept of the "symbiont" and the "symbiome". Symbionts are the units of symbiogenesis, and the symbiome is the outcome of symbiogenesis. In cultural evolution, scientific theories or cultural ideas, for example, might be either symbionts or symbiomes, if they evolved out of the synthesis of different streams of thought. These notions can therefore complement Hull's "interactors" and Dawkins' "replicators".

6.4 Applying Universal Symbiogenesis

Where can we apply this universal scheme of symbiogenesis? It will be evinced that universal symbiogenesis can be distinguished, at the very least, in the evolution of viruses, plants, languages and cultures.

6.4.1 Universal Symbiogenesis and Hybridization

Under certain conditions, a symbiotic relation between organisms can become hereditary. The botanist Joshua Lederberg was the first to notice that when such "hereditary symbiosis" occurs, it highly resembles plant hybridization. It resembles hybridization in that sense that phylogenetically distinct genomes are blended into one organism and this process is often irreversible (Sapp 2003: 244; Carrapiço 2006). Symbiosis can thus be regarded as a type of hybridization, or hybridization can be understood as a special kind of symbiosis because both mechanisms can introduce evolutionary novelty through the combination of different phylogenetic lineages.

Traditionally, especially zoologists tended to classify hybrids as exceptional sports of nature that are irrelevant in the explanation of the origin of new species. Examples such as the donkey or mule allowed them to argue that the crossings of lineages leads to infertility and when infertility occurs, evolutionary lineages end. Hence, no further evolution can occur. But today, due to important progress that is being made within both botany as well as zoology, hybridization is now considered a major creative force (Ryan 2006). Nowadays, we know that the crossings of lineages often allows for the introduction of novel features. And, most importantly, hybridization does not necessarily lead to infertility and the subsequent ending of a stock. In fact, most hybrids are able to reproduce just fine.

These new insights pose an interesting challenge to Neodarwinian, selectionist thought. The latter traditionally understands the evolution of a new species to occur vertically, through a branching off from a common stock. Mayr's (1997) biological species concept argues that the impossibility to produce fertile offspring is the key criteria to recognize the evolutionary emergence of a new species. In fact, the term speciation, i.e. the formation of new species, is nowadays considered to be synonymous to branching or splitting off. But Mayr's characterization of speciation also immediately implies that members of the same species that belong to opposite sexes are able to produce fertile offspring. Thus the possibility for different organisms to cross horizontally or to "hybridize", cannot allow for new species to emerge. On the contrary, the possibility for organisms to exchange genetic material successfully is considered the criterion for including these individuals into the same species.

Hybridization, understood as an evolutionary mechanism that can allow for the introduction of new species, on the contrary, takes crossings as the criterion for speciation. Here, the different mergings are what allows for the distinction into species.

Speciation understood as the introduction of novel species by means of vertical splitting processes; or speciation understood as the introduction of novel species by means of horizontal hybridization processes, are mutually exclusive concepts. They pose a challenge to evolutionary scholars who in the future need to figure out how both can explain the evolution of life.

Nonetheless, the point being made here is that hybridization can be considered a type of universal symbiogenesis: the horizontal mergings of different entities can rapidly result into the emergence of new variants if not species altogether. And this process is often irreversible. The jargon of universal symbiogenesis can thus be applied to both the origin of eukaryotes as well as the origin of hybrid plants and animal species, thereby proving its qualities as a universal methodology.

6.4.2 *Universal Symbiogenesis and Viruses*

Several authors (Gontier 2006b: 204–6; Roosinck 2005; Ryan 2002, 2004, 2006; Sapp 2003; Villarreal 2004) have recently proposed that the scientific jargon of symbiogenesis can also be extended to the domain of virology.

Viruses are often considered to parasite their hosts, but the virus-host interaction can quite often be viewed as a symbiotic union. Viruses can introduce their genetic material into genomes of somatic cells and also into germ cells. These genetic exchanges occur horizontally, during the ontogeny of their host, and when the germ cells are infected, the newly introduced viral genes can possibly be transmitted during phylogeny. Especially such germ line transmission of viral genes, also called "viral colonization" by Villarreal (2004: 315) can possibly lead to the introduction of novel features in the host (Villarreal 2004: 296), if not lead to the evolution of new species altogether.

In line with the universal symbiogenetic template, viruses can therefore be considered as agents that, through their mergings with other agents, allow for the rapid emergence of novel features.

Evidence that viral infection of host germ cells can lead to the introduction of new features comes from ERVs, endogenous retroviruses. Retroviruses are viruses such as the HIV virus. Endogenous retroviruses are retroviruses that have become part of the genome. The non-coding regions of the genomes of all vertebrates (the so-called Junk DNA) all contain parts and pieces of ERVs (Ryan 2004: 560, Villarreal 2004: 297–298). In humans, they even make up half of the genome, where "*they replicate in Mendelian Fashion, as an integrated part of the sexual reproduction of the host, to inhabit the genome of all future generations*" (Ryan 2004: 560).

Because of their high occurrence in the vertebrate genome, evolutionary geneticists are increasingly willing to attribute to these ERVs a major role within evolution. Ryan (2002) pictures our past to be characterized by "plague culling": he envisions vertebrate species to have been frequently plagued by exogenous retroviruses. Such epidemics most probably resulted in the weeping out of entire species. But when the genotypes of host individuals were able to cope with the constant presence of the viruses, the retroviruses became endogenous, and a new symbiotic union arose (Ryan 2004: 561).

Eventually, some of these originally quiet genes might have become translated and their expression might influence developmental pathways and eventually give rise to speciation. Ledeberg (Sapp 2003: 243, and 2004: 1048) already assumed that such "infective heredity" might be possible and molecular genetics can nowadays easily test these hypotheses.

Because symbiogenesis can be expanded to include processes in virology, Ryan (2002: 117) redefines symbiogenesis as follows:

> Symbiogenesis is evolutionary change arising from the interaction of different species. It takes two major forms: endosymbiosis, in which the interaction is at the level of the genomes, and exosymbiosis, in which the interaction may be behavioural or involve the sharing of metabolites, including gene-coded products.

This expansion of symbiogenesis into the viral realm again demonstrates the potential of symbiogenesis to provide a transdisciplinary, universal jargon if not a methodology to study certain horizontal evolutionary events.

Villarreal (2004: 304) even goes on to argue that also the eukaryotic nucleus has a viral origin. Margulis' (1999) SET-theory hypothesizes that the nucleus is the result of the symbiotic merger of different prokaryotes. Villarreal (2004: 304–5) disagrees that it were prokaryotes that engaged in the symbiotic merger. He compared the genes of all prokaryotic organisms and found that there are 324 genes held in common. However, none of these genes are involved in DNA replication. The genes responsible for the latter thus must come from a different source. Villarreal (2004) and Villarreal and Defilipps (2000) take viruses to be the most likely candidates to have donated the genes responsible for the formation of the eukaryotic nucleus. In his account, viruses are "gene-creating machines". Around 80% of the genes that are found in viruses are not found in any other pro- or eukaryote. They are thus unique to viruses. According to Villarreal it were viruses that engaged in symbiogenetic relations, and this resulted in the origin of the nucleus.

Finally, a very new and highly promising application of symbiogenetic jargon can be found in medicine. Today, physicians are genetically modifying viruses or bacteria which they subsequently induce into patients as a type of treatment for certain illnesses (Ryan, personal communication). In this regard, the process of symbiogenesis is artificially applied.

6.4.3 Universal Symbiogenesis and Culture

So far we have seen how universal symbiogenesis can fruitfully be applied within the most important fields of biology, including (evolutionary) zoology, botany, virology, and exobiology. But can the jargon of symbiogenesis also be expanded to extra-biological fields such as culture? Here it will be proven that this can indeed be done. The reason for this is that also in culture, the blending and crossing of different cultural artefacts or ideas occurs very frequently.

6.4.3.1 Cultural Anthropology and the Nature Culture Divide

The divide between nature and culture is an old discussion that has already been debated upon by the ancient Greeks if not by more ancient civilizations. The contemporary version of the divide dates back to the nineteenth and twentieth century, when the rising fields of anthropology and sociology would set themselves off against evolutionary biology (Eriksen and Nielsen 2001: 16–36; Ingold 1986: 30–73).

The dichotomization between the fields was mainly the result of the evolutionists' turn to social Darwinism, eugenetics and historicism (e.g. Darwin 1871, 1872; Spencer 1976, 1978; Haeckel 1883, 1912; and Galton 1909). The latter falsely assumed that the human species could be divided into different "races". They postulated the existence of developmental laws of nature and culture and argued that not all "ethniés" were equally evolved. In other words, they argued that cultural diversity was the result of biological, racial diversity.

Anthropologists such as Boas (1924, 1930, 1932, 1962), Kroeber (1963), Malinowski (1944) and Radcliffe-Brown (1957) broke with this tradition and endorsed the idea of psychic unity: they argued that all humans were equally evolved both at a biological as well as a cultural level. Rather than argue that biology determined the level of culture or civilization one finds oneself in, they argued that there is not one biologically determined culture. Rather, there exist a multitude of unrelated cultures. All are particular and contingent upon the society one lives in. Culture, in this regard, is what is learned or nurtured, while nature is what is innate.

Because of the equation of culture with nurture, and nature with innateness and heredity of this innateness, the idea grew that nature and culture evolve according to radically different processes. Nature is classically argued to evolve through descent with modification (as the Neodarwinian doctrine describes), and thus biological evolution is considered to be vertical. The transmission of different cultures however is mostly considered to be a horizontal event, since people are not only encultured by their parents, but also by their siblings, peers, caregivers, teachers, etc. To explain these kinds of horizontal transmission processes, anthropologists use concepts such as acculturation, diffusion or integration.

Furthermore, when evolutionary biologists identified the genes as the carriers of hereditary information, they also found out that the genes that encode for features are more often than not passed on unchanged. In this regard, genes are very rigid and integer. Cultural traits, on the other hand, are not unchanging, rigid entities. Cultures constantly change because new ideas (e.g. the risks of global warming) or artefacts (e.g. cell phones, iPods, radars) are introduced that change our way of living. Cultural traits are also not integer. The multitude of ideas or artefacts that make up a culture or a cultural individual are not always invented in that specific culture by that specific individual. In fact, the majority of cultures is made up of ideas and artefacts that originated in other cultures. These traits are acquired through trade, diffusion, politics, warfare, cultural hegemony, etc. In this regard, Kroeber, one of the founding fathers of cultural anthropology, noted the following:

> “[T]he great part of the content of every culture is probably of foreign origin, although assimilated into a whole that works more or less coherently and is felt as a unit. However diversified or specialized a culture grows in its development, it can therefore always largely retrace its course; and it does normally do so, by absorbing more generalized content from other cultures, and thereby not only assimilating to them but to the totality or the average of human cultures.” (Kroeber 1963: 67–8)

Communism, for example, arose in the West but is now the political system of Cuba and China; Russia became capitalistic; Christmas is celebrated on the ancient Roman fest of light because ancient scholars conflated the birth of Jesus with pagan celebration of the re-birth of light; nachos with salsa and guacamole are becoming the favourite appetizers in the North of Europe; western music is played on mp3 players that are made in China; the dress style of the 1960s and 1970s is nowadays fashionable again and has been dubbed “vintage”, etc.

Culture does not follow a linear means of transmission that is typical of evolution by means of natural selection. Rather, grandchildren can teach their grandparents how to use the computer, Americans can teach the Japanese how to dance the foxtrot, or a contemporary philosopher can learn about logic by reading the works of Aristotle.

To illustrate these kinds of cultural transmission processes, Kroeber argued that cultural evolution needs to be depicted by a different tree than the one that depicts evolution by means of natural selection (Fig. 6.1).

Cultural evolution is not merely characterized by a branching off of lineages from one common stock as is the case with evolution by means of natural selection. Rather, lineages can cross and blend and a variety of common ancestors can be identified for these different lineages. Kroeber:

> Once the genetic diversification or ‘evolution’ has gone beyond a certain quite narrow degree, there is no more possibility of reversal and assimilation. By contrast, cultures can blend to almost any degree and not only thrive but perpetuate themselves. (Kroeber 1963: 67–8)

Cultures absorb new elements all the time and these mergings can even be traced, to a certain extent, because the cultural influences leave traces (in written texts or traditions). Furthermore, contrary to evolution by means of natural selection, the transmission of culture can happen very quickly. Almost anywhere in the world, old photo cameras have been replaced by digital ones, flat screens are part of many households in the West, youth everywhere is on Facebook, My Space, Twitter or You Tube and none of this existed 10 or sometimes even 5 years ago.

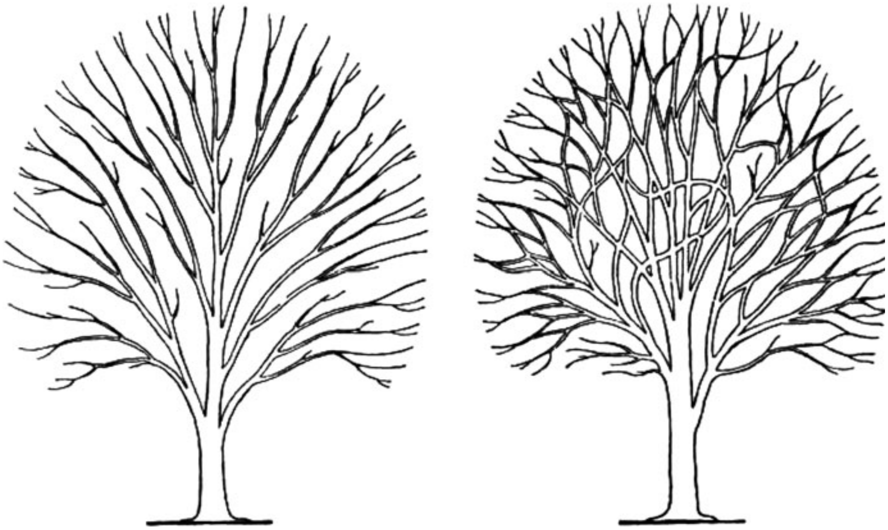


Fig. 6.1 Kroeber (1963: 68) depicted the tree of life to be a process of branching, as is dictated by the theory of evolution by means of natural selection (on the *left*); while he portrayed cultural evolution as a tree that contains branches that do not merely split, but also merge (on the *right*)

Regarding his cultural tree, Kroeber (1965: 68) wrote the following:

> … the course of organic evolution can be portrayed properly as a tree of life, as Darwin has called it, with trunk, limbs, branches, and twigs. The course of the development of human culture in history cannot be so described, even metaphorically. There is a constant branching-out, but the branches also grow together again, wholly or partially, all the time. Culture diverges, but it syncretizes and anastomoses too.

Given the above outline on horizontal evolution within different fields of evolutionary biology, the parallel with evolution through symbiogenesis and the cultural transmission model depicted by Kroeber is obvious. Symbiogenesis too explains how different lineages can cross, merge and rapidly lead to new entities. At the time that Kroeber and Boas were formulating their ideas on culture, de Bary, Frank and Merezhkowsky were also writing on symbiogenesis (for an excellent overview see Sapp et al. 2002). Unfortunately, the latter's ideas did not reach mainstream evolutionary biology let alone cultural anthropology.

What happened in anthropology is therefore somewhat unfortunate. Not being able to depict let alone explain cultural evolution by means of natural selection, anthropologists gave up an evolutionary biological approach to culture altogether. For many years, anthropologists would argue that cultural evolution is a process so different from the evolution of life (equated with evolution by means of natural selection), that methodologies other than evolutionary biological ones need to be applied to the subject. In fact, many anthropologists even refused to use terms such as "evolution" altogether and preferred terms such as "cultural transmission" or "diffusion" instead (Borgerhoff Mulder et al. 2006). In this respect, two different scientific languages emerged, one for the natural sciences, and one for the humanities. Nonetheless, acculturation or diffusion can also be seen as means or vehicles of universal symbiogenesis, just as symbiosis is one of the vehicles for endosymbiogenesis.

6.4.3.2 Hybridization Models of Culture and Language

With the success of the Modern Synthesis and the rise of Post-Neodarwinian, sociobiological theory (Dawkins 1976; Williams 1966), a minority of anthropologists turned to biology in order to explain certain cultural phenomena. Eventually, evolutionary psychology (Dunbar et al. 1995) would take on the investigation of culture from within a selectionist framework. But evolutionary psychologists, as the name implies, are mostly found in the psychology department.

When social and cultural anthropologists make use of biological thought to explain culture, they by and large still turn to fields other than evolutionary psychology. The fact that different cultures live in close contact with one another for instance has lead to the introduction of cultural ecology (Ingold 1986: 40), an approach that has loose parallels with the ecological approach within biology. But both in biology as well as in anthropology, an ecological approach does not immediately imply an evolutionary approach.

Furthermore, processes such as culture contact, language mixing or language borrowing are often compared to hybridization processes in plants. Anthropologists interested in multiculturalism, for instance, launch concepts such as cultural hybrids to explain transnationality, creolization, acculturation, deculturation or ethnic shopping (Chavez 2006; Hannerz 1980, 1992, 2002; Pinxten and De Munter 2006). With these notions they want to capture the rapid flow of culture from one person to the next, within different generations. Individuals or cultural groups constantly interact with one another and are interconnected, controlled, absorbed, or rejected through economy, life style, religion and ideology. The notions of symbiont and symbiome can be applied as a universal delineator that captures all these cultural concepts; and creolization, deculturation or ethnic shopping can be regarded as vehicles of universal symbiogenesis.

Hybridization models and ecological approaches are also found within sociolinguistics to explain phenomena such as language variation, language contact, language mixing and language borrowing. Mufwene (2002, 2005) has developed an ecological approach to language evolution/variation, based upon the neo-Darwinian population geneticist approach. Although the latter would imply that he emphasizes selection and thus speciation, he actually emphasizes the contact that exists between individual idiolects, and their possibility to exchange information and merge, similar to the reproduction process in biology where members of the same species and of the opposite sex can exchange genes. Rather than focus on the vertical selection phase, Mufwene thus uses the horizontal moment of genetic recombination during reproduction to draw analogies between language evolution and the evolution of life.

Croft (2000, 2002) actually uses an evolutionary epistemological model to explain the evolution/variation of languages. More specifically, he makes use of Hull's (1988) replication variation and environmental interaction scheme to explain language variation. However, he also recognizes that the latter approach is insufficient to explain all phenomena in language variation. Rather, language mixing and language borrowing is better explained by using metaphors of hybridization, which Croft dubs the "*plantish approach*".

Both Croft and Mufwene are working within the field of historical linguistics, also known as sociolinguistics or diachronic linguistics. The latter traditionally investigate language variation and the historical diversification of languages, but they do not a priori work within an overall evolutionary biological framework. And although both Croft and Mufwene make use of the term language evolution, they do not speak of language evolution in the same sense as evolutionary linguistics (Hurford et al. 1998) talk about it. The latter are much more inspired by evolutionary psychological, selectionist approaches. Nonetheless, both Croft as well as Mufwene argue that their models can also be applied within the field of evolutionary linguistics where they can help in the study of the actual evolution of languages.

However useful, hybridization models, both in culture as well as in linguistics, can be misleading. For one, hybridization is often not regarded as an evolutionary process, but as we have already demonstrated above, the latter is false. Hybridization can indeed be considered a creative force in evolution, and a specific example of symbiogenesis.

But cultures are not like animal or plants species when they hybridize. Hybridization can be a rather rigid process, that always implies the emergence of a new generation. Languages and cultures, or individual speakers can always absorb new elements without having to produce a new generation. Furthermore, individuals can easily undo themselves of acquired cultural traits or dialects, while animal and plant hybrids cannot, the hybridization is often permanent.

Therefore, it is better to use the jargon of universal symbiogenesis to describe the horizontal evolution of language and culture and therefore to bridge the gap between the natural and life sciences, as Neurath already pleaded for. Biological individuals, their languages and their cultures, all are chimeras, entities that are stitched and patched together and that form new wholes with behaviours not seen in the individual parts.

Socialism for example is a symbiont or symbiome of communism, Marxism, capitalism, and enlightenment theory. A computer is a combination of electricity, light, plastic, steel, typewriters and television screens (Margulis and Sagan 2002: 172). None of them evolved to become part of a modern day computer, but at one point in history, they all got symbiogenetically combined, leading to a new entity with specific behaviour.

The universal symbiogenetic jargon can thus be expanded to the humanities and provide a universal methodology to examine the above described horizontal forms of cultural evolution. And also in language evolution, the jargon can be applied successfully. Creoles for example can be understood as chimeras that combine the host lexifier and symbiont substrate language, or vice versa.

In sum, the universal symbiogenetic account can also be put to use in the humanities and here it can provide a complementary view to selectionist speciation models that all too often overemphasize selfishness and competition. The merit of universal symbiogenesis is that it allows for the introduction and even for the recognition of interaction, cooperation and exchange (especially in this regard, see Speidel 2000).

6.5 Conclusion

Traditionally, EE was understood to be a discipline that studied cognition and classical philosophical problems from within evolutionary theory. As such, it did not differ much from Quine's naturalized epistemology. However, through the years, EE has become a discipline that not only tackles philosophical problems from within natural selection theory. It also studies phenomena other than cognition and deals with problems other than knowledge. Contrary to traditional EE that focussed exclusively on natural selection as the evolutionary mechanism that allows for the biologizing of epistemology, scholars working within new EE are also looking at other evolutionary theories to study traditional epistemological questions.

The merit of all these approaches however also make it necessary to investigate whether it is possible to epistemologize evolutionary theories other than natural selection. Applying the tenets of new evolutionary epistemology to all known evolutionary mechanisms will provide us with more methodologies to examine evolution at all ranks of life.

In this article, it was specifically investigated how symbiogenesis can be universalized and how it can provide a methodology in both the life and human sciences.

It was demonstrated that symbiogenesis can be universalized and in its universal form, it can include at minimum the epidemiology of viruses, hybridization, cultural and linguistic evolution. Universal symbiogenesis even has potential in medical applications. Moreover, the well-used notions of *symbiont* and *symbiome* can be applied as universal concepts that can complement Dawkins' replicators and Hull's interactors.

The enormous potential of an evolutionary view based on symbiogenesis is yet to be felt in many extra-biological fields and even within zoological-centred evolutionary biology. The universal symbiogenetic formula presented in this article hopes to contribute in a positive way in making the importance of symbiogenesis known in these fields. In the wake of Neurath, it will introduce another language if not methodology, that will bring the natural and the life sciences together.

Acknowledgements Sincere thanks to Frank Ryan and Francisco Carrapiço for providing helpful comments on parts of the script and to the editors for their excellent work on the volume.

References

Barrett, L., R. Dunbar, and J. Lycett. 2002. *Human evolutionary psychology*. Hampshire: Palgrave.

Blackmore, S. 1999. *The meme machine, with a foreword of Richard Dawkins*. Oxford: Oxford University Press.

Boas, F. 1924. Evolution or diffusion. In Boas, F. 1940. *Race, language and culture*, 290–294. New York: Macmillan.

Boas, F. 1930. Some problems of methodology in the social sciences. In Boas, F. 1940. *Race, language and culture*, 260–269. New York: Macmillan

Boas, F. 1932. The aims of anthropological research. In Boas, F. 1940. *Race, language and culture*, 243–259. New York: Macmillan.

Boas, F. 1962. (1928) *Anthropology and modern life*. New York: W.W. Norton and company.

Borgerhoff Mulder, M., C.L. Nunn, and M.C. Towner. 2006. Cultural macroevolution and the transmission of traits. *Evolutionary Anthropology* 15: 52–64.

Bradie, M. 1986. Assessing evolutionary epistemology. *Biology and Philosophy* 1: 401–459.

Bradie, M. and W. Harms. 2001. Evolutionary epistemology. *The Stanford Encyclopaedia of Philosophy* 1–13. http://www.compilerPress.atfreeweb.com/Anno%20Bradie%20&%20Harm%20Evol%20Epist.htm.

Brandon, R.N. 1982. The levels of selection. In *Genes, organisms, populations: Controversies over the units of selection*, ed. R.N. Brandon and R.M. Burian, 1984, 133–139. Cambridge: Massachusetts Institute of Technology.

Brandon, R.N. and R.M. Burian, R.M. eds. 1984. *Genes, organisms, populations: controversies over the units of selection.* Cambridge, MA: MIT.

Callebaut, W., and R. Pinxten. 1987. Evolutionary epistemology today: Converging views from philosophy, the natural and social sciences. In *Evolutionary epistemology: A multiparadigm program with a complete evolutionary epistemology bibliography*, ed. W. Callebaut and R. Pinxten, 3–55. Dordrecht: Reidel.

Campbell, D.T. 1974. Evolutionary epistemology. In *The philosophy of Karl Popper*, vol. 1, ed. P.A. Schlipp, 413–459. La Salle: Open Court Publishing Co.

Campbell, D.T. 1997. From evolutionary epistemology via selection theory to a sociology of scientific validity: Edited by Cecilia Heyes and Barbara Frankel. *Evolution and Cognition* 3(1): 5–38.

Carrapiço, F. 2006. The origin of life and the mechanisms of biological evolution. *Proceedings of SPIE* 6309: 630900-1-5.

Carrapiço, F., and T. Rodrigues. 2005. Symbiogenesis and the early evolution of life. *Proceedings of SPIE* 5906: 59060R-1-4.

Changeaux, J.P. 1985. *Neuronal man: The biology of mind*. New York: Oxford University Press.

Chavez, L.R. 2006. Culture change and cultural reproduction: Lessons from research on transnational migration. In *Globalization and change in fifteen cultures: Born in one world and living in another*, ed. J. Stockard and G. Spindler. Belmont: Thomson-Wadsworth.

Croft, W. 2000. *Explaining language change: An evolutionary approach*. Essex: Pearson.

Croft, W. 2002. The Darwinization of linguistics. *Selection* 3(1): 75–91.

Cziko, G. 1995. *Without miracles: Universal selection theory and the second Darwinian revolution*. Cambridge: Massachusetts Institute of Technology.

Darwin, C. 1871. *The descent of man, and selection in relation to sex*, vol. 2. London: John Murray. http://darwin-online.org.uk/content/frameset?itemID=F937.1&viewtype=text&pageseq=1&http://darwin-online.org.uk/content/frameset?itemID=F937.2&viewtype=text&pageseq=1.

Darwin, C. 1872. *The expression of the emotions in man and animals*. London: John Murray. http://darwin-online.org.uk/content/frameset?itemID=F1142&viewtype=side&pageseq=1.

Dawkins, R. 1976. *The selfish gene*. New York: Oxford University Press.

Dawkins, R. 1982. Replicators and vehicles. In *Genes, organisms, populations: Controversies over the units of selection*, ed. R. N. Brandon and R. M. Burian, 1984, 161–179. Cambridge: Massachusetts Institute of Technology.

Dawkins, R. 1983. Universal Darwinism. In *The philosophy of biology*, ed. D.L. Hull and M. Ruse, 15–35. New York: Oxford University Press.

Dunbar, R.I., N.D. Duncan, and D. Nettle. 1995. Size and structure of freely forming conversational groups. *Human Nature* 6: 67–78.

Dyson, F. 1988. *Infinite in all directions*. London: Penguin.

Dyson, F. 1998. The evolution of science. In *Evolution: Society, science and the universe*, ed. A.C. Fabian, 118–35. Cambridge: Cambridge University Press.

Dyson, F. 1999. *Origins of life: Revised edition*. Cambridge: Cambridge University Press.

Edelman, G. 1987. *Neural Darwinism: The theory of neuronal group selection*. New York: Basic Books.

Eigen, M. 1996. *Steps towards life: A perspective on evolution*. New York: Oxford University Press.

Eigen, M., and P. Schuster. 1977. The hypercycle: A principle of natural self-organisation. Part A: Emergence of the hypercycle. *Naturwissenschaften* 64: 541–65.

Eriksen, T.H., and F.S. Nielsen. 2001. *A history of anthropology*. London: Pluto Press.

Fox, S.W., and K. Dose. 1972. *Molecular evolution and the origin of life*. San Francisco: W.H. Freeman & Co.

Galton, F. 1909. *Essays in eugenetics*. London: The eugenetics education society.

Gilbert, W. 1986. The RNA world. *Nature* 319: 618.

Gontier, N. 2006a. Introduction to evolutionary epistemology, language and culture. In *Evolutionary epistemology, language and culture*, ed. N. Gontier, J.P. Van Bendegem, and D. Aerts, 1–29. Dordrecht: Springer.

Gontier, N. 2006b. Evolutionary epistemology and the origin and evolution of language: Taking symbiogenesis seriously. In *Evolutionary epistemology, language and culture*, ed. N. Gontier, J.P. Van Bendegem, and D. Aerts, 195–226. Dordrecht: Springer.

Gontier, N. 2006c. Evolutionary epistemology. In *The Internet encyclopaedia of philosophy.*http://www.iep.utm.edu/e/evo-epis.htm.

Gontier, N. 2007 Universal symbiogenesis: An alternative to universal selectionist accounts of evolution. *Symbiosis* 44: 167–181.

Gould, S.J., and R.C. Lewontin. 1979. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist program. *Proceedings of the Royal Society of London, B* 205: 581–589.

Haeckel, E. 1883. (1868) *The history of creation*, vol. 1 (trans: Lankester, E.R. and Kegan, P.). London: Trench & Co.

Haeckel, E. 1912. *The evolution of man, Vol. I: Human embryology or ontogeny*, vol. I. London: Watts & Co [Translated from the fifth enlarged edition by McCabe, J.].

Hahlweg, Kai, and C.A. Hooker. 1989. A generalized framework for evolutionary processes. In *Issues in evolutionary epistemology*, ed. Kai Hahlweg and C.A. Hooker, 79–100. New York: State University of New York Press.

Hannerz, U. 1980. *Exploring the city: Inquiries toward an urban anthropology*. New York: Colombia University Press.

Hannerz, U. 1992. *Cultural complexity: Studies in the social organization of meaning*. New York: Colombia University Press.

Hannerz, U. 2002. Flows, boundaries and hybrids: Keywords in transnational anthropology, 1–25. Stockholm University, Department of Social Anthropology, Unpublished manuscript http://www.transcomm.ox.ac.uk/working%20papers/hannerz.pdf#search=%22%22flows%2C%20boundaries%20and%20hybrids%22. (Translation of Hannerz, U. 1997. Fluxos, fronteiras, hîbridos: palavras-chave da antropologia transnacional. *Mana* (Rio de Janeiro), 3(1): 7–39.)

Hull, D.L. 1980. Individuality and selection. *Annual Review of Ecology and Systematics* II: 311–332.

Hull, D.L. 1981. Units of evolution. In *Genes, organisms, populations: controversies over the units of selection*, ed. R.N. Brandon and R.M. Burian, 1984, 142–159. Cambridge: Massachusetts Institute of Technology.

Hull, D.L. 1988. *Science as a process: An evolutionary account of the social and conceptual development of science*. Chicago: The University of Chicago Press.

Hurford, J., M. Studdert-Kennedy, and C. Knight (eds.). 1998. *Approaches to the evolution of language*. Cambridge: Cambridge University Press.

Ingold, T. 1986. *Evolution and social life*. Cambridge: Cambridge University Press.

Kroeber, A.L. 1963. *Anthropology: Culture patterns and processes*. New York: Harbinger Books [First edition 1923].

Lewontin, R. 1970. The levels of selection. *Annual Review of Ecology and Systematics* 1: 1–18.

Lewontin, R. 2000. *The triple helix: Gene, organism and environment*. Cambridge: Harvard University Press.

Lorenz, Konrad. 1941. Kants Lehre vom Apriorischen im Lichte gegenwärtiger Biologie. *Blätter für Deutsche Philosophie* 15: 94–125 (English translation in Plotkin, Henry C., *op. cit.* 121–143.).

Malinowski, B. 1944. *A scientific theory of culture and other essays, with a preface by Huntington Cairns*. Chapel Hill: University of North Carolina Press.

Margulis, L. 1999. *The symbiotic planet, a new look at evolution*. Phoenix/London: Orion books.

Margulis, L., and M.F. Dolan. 2002. *Early life: Evolution on the Pre-Cambrian earth*, 2nd ed. Sudbury: Jones and Bartlett.

Margulis, L., and D. Sagan. 2000. *What is life?* Berkeley: University of California Press.

Margulis, L., and D. Sagan. 2002. *Acquiring genomes: A theory of the origin of species*. New York: Basic Books.

Mayr, E. 1997. *Evolution and the diversity of life: Selected essays*. Harvard: Harvard University Press.

Mufwene, S. 2002. What do Creoles and Pidgins tell us about the evolution of language? Unpublished manuscript available at: http://humanities.uchicago.edu/faculty/mufwene/CREOLES-LGEVOLUTION-Revisions-1.pdf.

Mufwene, S. 2005. Language evolution: The population genetics way. In *Genes, languages and their evolution*, ed. G. Hauska, 30–52. Regensburg: Universitätsverlag Regensburg.

Munz, Peter. 2001 (1993). *Philosophical Darwinism: On the origin of knowledge by means of natural selection*. London: Routledge.

Neurath, Otto. 1936. *Une encyclopédie internationale de la science unitaire: actes du congrès international de philosophie scientifique – Sorbonne 1935* (trans: Symons, John and Alvarado, Ramon), unpublished. Paris: Hermann & Cie.

Oparin, A. 1955. *L'origine de la vie*. Moscou: Editions en langues étrangères.

Orgel, L.E. 1994. The origin of life on earth. *Scientific American* 271: 53–61.

Pinker, Steven, and Paul Bloom. 1990. Natural language and natural selection. *Behavioral and Bran Sciences* 13(4): 707–84.

Pinxten, R., and K. De Munter. 2006. *De culturele eeuw*. Antwerp: Houtekiet [The cultural century].

Plotkin, H. 1995. *Darwin machines and the nature of knowledge: Concerning adaptations, instinct and the evolution of intelligence*. London: Penguin.

Popper, Karl. 1963. *Conjectures and refutations*. London: Routledge & Kegan Paul.

Radcliffe-Brown, A.R. 1957. *A natural science of society*. Glencoe: The Free Press.

Riedl, R. 1984. *Biology of knowledge: The evolutionary basis of reason*. New York: Wiley.

Roosinck, M. 2005. Symbiosis versus competition in plant virus evolution. *Nature Reviews Microbiology* 3: 917–24.

Ryan, F. 2002. *Darwin's blind spot: Evolution beyond natural selection*. Boston: Houghton Mifflin Company.

Ryan, F. 2004. Human endogenous retroviruses in health and disease: A symbiotic perspective. *Journal of the Royal Society of Medicine* 97: 560–65.

Ryan, F. 2006. Genomic creativity and natural selection: A modern synthesis. *Biological Journal of the Linnean Society* 88: 655–672.

Sapp, J. 2003. *Genesis: The evolution of biology*. New York: Oxford University Press.

Sapp, J. 2004. The dynamics of symbiosis: An historical overview. *Canadian Journal of Botany* 82: 1046–56.

Sapp, J., F. Carrapiço, and M. Zolotonosov. 2002. Symbiogenesis: The hidden face of Constantin Merezhkowsky. *History and Philosophy of the Life Sciences* 24: 413–40.

Schwartz, J. 1999. *Sudden origins, fossils, genes and the emergence of species*. New York: Wiley.

Speidel, M. 2000. The parasitic host: Symbiosis contra Neo-Darwinism. *Pli, The Warwick Journal of Philosophy* 9: 119–38.

Spencer, H. 1976 (1897). *The principles of sociology*, 3 vols. New York: D. Appleton and Co.

Spencer, H. 1978 (1879). *The principles of ethics*, 2 vols. Indianapolis: Liberty Classics

Toulmin, Stephen. 1972. *Human understanding: The collective use and evolution of concepts*. Princeton: Princeton University Press.

Villarreal, L. 2004. Can viruses make us human? *Proceedings of the American Philosophical Society* 148(3): 296–323.

Villarreal, L., and V. Defilipps. 2000. A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *Journal of Virology* 74(15): 7079–7084.

Williams, G.C. 1966. *Adaptation and natural selection*. Princeton: Princeton University Press.

Wuketits, F.M. 1990. *Evolutionary epistemology and its implications for humankind*. New York: State University of New York Press.

Wuketits, Franz M. 2006. Evolutionary epistemology; The non-adaptationist approach. In *Evolutionary epistemology, language and culture: A non-adaptationist systems theoretical approach*, ed. Nathalie Gontier, Bendegem Van, Paul Jean, and Diederik Aerts, 33–46. Dordrecht: Springer.

Chapter 7
The Symbiotic Phenomenon in the Evolutive Context

Francisco Carrapiço

Abstract We live in a symbiotic world, and one of the key characteristics of the biological systems is to establish associations and connections with other organisms. This manifestation is one of life's main characteristics and its diversity. In a way, life has not established itself or developed to exist alone. Since the introduction of the symbiosis concept by Anton de Bary in 1878 and the new theoretical formulation on the field – symbiogenesis – by Constantin Mereschkowsky, in 1909, this domain of science has been a place of controversy and discussion. The symbiogenesis concept was a landmark for the development of further studies on biology and evolution, that the most remarkable example was the development of the Serial Endosymbiotic Theory (SET) by Lynn Margulis in 1967. Throughout the twentieth century, biologists have generally considered symbiosis as a curiosity. Its study fell largely outside the conceptual and technical framework of biology, and namely of neo-Darwinism. However, most living forms have symbiotic relationships with microorganisms, and so symbiosis seems to play a very important role in the origin, organization and life evolution. In this sense, evolutionary changes can be explained by an integrated and synergistic cooperation between organisms, in which symbiosis acts, not as an exception, but as the dominant rule in nature. Without denying many of the Darwinist principles, the worse thing we could do within the study of the process of evolution would be to mix up or limit evolution to the Darwinist or neo-Darwinist perspective. Other evolutionist approaches exist and it is necessary for them to be deepened and discussed within biological and social sciences. In this sense, we would like to bring a set of principles and data that could be integrated in a Symbiogenic Theory of Evolution that could contribute towards a new epistemological approach of the symbiotic phenomenon in the evolutive

---

F. Carrapiço (✉)
University of Lisbon, Faculty of Sciences, Department of Plant Biology, Centre for Environmental Biology and Centre for Philosophy of Sciences, Campo Grande, 1749-016 Lisbon, Portugal
e-mail: fcarrapico@fc.ul.pt

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology, and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_7,

context. This, in our point of view, could be the beginning of a new paradigm in science that rests almost unexplored.

Keywords New paradigm • Symbiogenesis • Symbiosis evolution

7.1 Introduction

Symbiosis has frequently been considered as a biological curiosity, and not as a solid scientific concept, namely in the scope of evolution. Nevertheless, symbiosis is a widespread phenomenon, with great biological relevance, that has a fundamental role in the origin, organization and evolution of life. Despite this fact, it has not received proper attention either from the scientific community, or from the university and high school curricula. This situation reflects itself in an interpretative reality of the biological evolution, based on two classical scientific theories: Darwinism and neo-Darwinism, the only ones that, traditionally, try to explain this process and where symbiosis is not adequately understood nor handled. For traditional evolutionist authors, symbiosis is nothing more than a residual aspect of the evolution problem, and its study fell largely outside the conceptual and traditional framework of biology, and namely of neo-Darwinism. Recent data, however, point to the opposite direction, showing that symbiosis is a factor of evolutive change, which cannot be included or adequately explained by the Theory of Modern Synthesis.

The symbiotic relations appear as dynamic relations that are not limited to the classical concepts of interspecific relations. Three main theoretical concepts are important to consider in the new approach on symbiogenic evolution. The first one is *symbiosis*, defined as "the living together of unlike named organisms", and introduced in 1878 by the German biologist Anton de Bary (De Bary 1878). Associated with this concept, is *symbiogenesis*, developed by the Russian biologist Constantin Mereschkowsky in 1909, as "the origin of organisms through the combination or association of two or more beings that enter in symbiosis", and based on the role of cyanobacteria in the origin of chloroplasts in plants (Mereschkowsky 1905, 1909). The third concept, presented by the American biologist Lynn Margulis in 1967 – the *Serial Endosymbiotic Theory* (SET) –, was a remarkable contribution to the rehabilitation and development of the symbiogenic ideas applied to the cellular world, explaining in an elegant way the transition bridge between the prokaryotic and the eukaryotic levels of the biological organization (Sagan 1967). This reality demonstrates the need for conceptual changes in the traditional vision that has been passed on to the organism's structures and functions, whose profound consequences for the biological, medical and social domain have remained practically unchanged.

7.2 Questions for the Twenty-First Century

One of the main concepts Charles Darwin contributed to change radically was the idea of the constancy of species, which allowed for the development of the theory of common descent and also challenged the natural theology principles that had ruled natural science for centuries. For natural theology, the order in nature was the convincing proof of a supreme being that could explain the harmony and the purpose of the creation. For this reason, we can understand the difficulties and the resistance which these new ideas were met when they were introduced to the general society (Mayr 1982).

Evolution is a complementary process of divergence and integration. Divergence in the production of new life forms, and integration when entities join to form new ones (Sapp 2003). In this context, evolution is a dynamic process that evolves and responds not in the sense of perfection or progress, but in the sense of adaptation to new conditions. We can ask in what way symbiosis can be involved or associated to evolution. In our point of view, symbiosis is the way through which the acquisition of new genomes and new metabolic capacities occur, making possible the evolutive construction of biological organisms. As it was referred by Joshua Lederberg (Lederberg 1952), endosymbiosis is comparable to hybridization, allowing for the introduction of philogenetically distinct genomes into associations of organisms with its own characteristics. We believe that it is held to have also played a central role in the pre-biotic evolution, in the emergency and evolution of eukaryotes, in the origin of land plants, and in a myriad of adaptive evolutionary innovations (Sapp 2003; Carrapiço et al. 2007). It is at the basis of important ecosystems from deep-sea vents to the most biodiverse communities on Earth, such as rainforests and coral reefs (Sapp 2003, 2004). The development of these new evolutive characteristics by the associated organisms is inconsistent with the main tenets of neo-Darwinism, but represents, in our opinion, the main rule in nature, and the main evolutionary mechanism in the establishment and maintenance of biomes, as well as the foundation of biodiversity.

Questions such as the following ones should have a clear and scientific answer and should not be put aside simply because they do not correspond to the mainstream discourse. Why is the symbiotic phenomenon so widespread in nature? Why and how does it perpetuate itself through time? What is its degree of importance for the intervening organisms? Why do living beings present structural, physiological and/or behavioural changes so well adapted to this relationship? What is the role of the symbiosis in the evolutionary process? In this sense, symbiogenesis should be understood as an evolutive mechanism and symbiosis as the vehicle, through which that mechanism unfolds (Chapman and Margulis 1998; Carrapiço et al. 2007). This fact represents a point of view different from that sustained by the Modern Synthesis, and opens new approaches for new models to understand the evolutive process.

7.3 Symbiogenic Revolution

One of the main characteristics of biological systems is the establishment of associations and connections with other organisms, thereby creating diversity and varied combinations of its forms, which are prodigious novelty generators. Thus, once established, life did not stop evolving or remained a single form. In this context, symbiosis is an important factor in generating evolutionary change, giving rise rather suddenly to evolutionary novelty, based on the creation of new metabolic, anatomical and organismal characteristics (Margulis and Fester 1991). Symbiosis is also a robust phenomenon, offering many opportunities for the occupation of ecological niches otherwise unviable, and is highly relevant for survival purposes. One good example can be found in the organisms living in the hydrothermal vents ecosystems, which survive and evolve based on symbiotic associations (Sapp 2003; Carrapiço et al. 2007). Symbiogenesis creates new important selection units (symbiomes) arising through the integration of varied parts followed by progressive differentiation of the whole, providing a competitive advantage that goes beyond traditional neo-Darwinian selection. In this theory, evolution is a gradual process, essentially consisting of a natural selection conducted on minimal phenotypical variations, which are a product of genetic and chromosome exchange. However, we think that natural selection acts not only in the minimal phenotypical variations, but also in variations resulting from different symbiotic adaptations.

Actually, and specially since 1859, evolution has been considered as the fundamental concept and organizing factor of modern Biology, as well as its structural pillar. Without denying many of the Darwinian principles, the worse thing we could do within the study of the process of evolution would be to mix up or limit evolution to the Darwinist or neo-Darwinist perspective. These points of view are mainly used to explain the biological evolution, contributing to the generalized belief that evolution could be explained by these two scientific theories. This led to the erroneous idea that Darwinism, or neo-Darwinism, are synonyms of biological evolution. Other evolutionary approaches exist and it is necessary for them to be deepened and discussed within biological sciences.

In this sense, we would like to bring to your attention a set of principles and data that could be integrated in a Symbiogenic Theory of Evolution. This theory includes Darwinist principles, but does not limit itself to the latter in its attempt to promote and explain the development, organization and evolution of the biological world. This approach does not present the cooperative perspective as the only leitmotif in its explanation of the biologic phenomena. As it was mentioned previously, considering symbiogenesis as an evolutive mechanism implies that evolution should be understood in a broader context, where symbiosis plays an essential role in the organization and structuring of the biological world. Consequently, the concept of symbiosis does not imply a strict compartimentation of interspecific relationships, thus it should be regarded as a continuous and dynamic process of different relations, such as mutualism, parasitism and commensalism. In this process, the acquisition of new genes through lateral transfer plays an important role. The same applies to the development of new metabolic capacities acquired by an organism from other organisms associated to it. The existence of mutual benefit should, however, not be considered, as the plus or common denominator of the symbiotic process, following the idea presented by Dubos and Kessler, in 1963, during the 1st International Conference on Symbiosis, in London, "the nutritional effects of symbiosis are not its most interesting manifestation. More remarkable is the fact that many symbiotic systems produce substances and structures that neither one of the two components produces when growing alone" (Dubos and Kessler 1963). This new approach cites the central role of interactions, in which a new entity emerges through incorporation of one existing entity into another. The theory involves horizontal mergers, which can be rapid, and are usually discontinuous, creating permanent and irreversible changes, the basis of evolutive novelty. The new entity can then evolve vertically, but this is always preceded by a horizontal merger between two or more entities.

In ecological terms, each plant and animal must be considered as "superorganism"-symbiome, which includes their own genes, those of cellular organelles (mitochondria and, or chloroplasts), as well as the genetic information of symbiont bacteria and virus living within the organism (Sapp 2003). This is also important to be considered when we look to the fitness and how we validate it in terms of symbiotic prevalence. It goes beyond the reproductive view of each individual and reinforces the ecological behaviour of the symbiotic system as a whole (Bouchard 2007). However, for many biologists, symbiosis is still considered as an exception among biological phenomena and not as a common rule in nature. The approach of the traditional evolutionist authors, in relation to this phenomenon, consists in considering it as nothing more than a residual aspect of the evolution problem. Recent data, however, point to the exact opposite direction, demonstrating that symbiosis is a factor of evolutive change, which cannot easily be included and explained in the framework of the neo-Darwinism theory, and moves the studies on evolution into the context of a post neo-Darwinian perspective (Carrapiço 2010).

7.4 Final Remarks

The understanding of the natural world is an important goal for mankind. For many scientists, the comprehension of nature is based on the assumption that "nature is competitive, and cooperation is a strange case that needs to be explained" (Speidel 2000). In this article we have presented several ideas about symbiosis in evolution, involving concepts that are not commonly taught or considered in current biology, but must be discussed in the domain of the philosophy of science. Evolution is usually taught as the result of mutations and genetic recombinations combined with natural selection, but most living forms have symbiotic relationships with microorganisms, and so symbiogenesis seems to play a very important role in the origin and life evolution. Symbiosis is an important support for the acquisition of new genomes and new metabolic capacities that drives living forms evolution. In this sense, evolutionary changes can be explained by an integrated synergistic cooperation between organisms, in which symbiosis acts, not as an exception, but rather as the rule in nature. Beginning with the eukaryotic cell formation, symbiogenesis appears to be the main evolutionary mechanism in the establishment and maintenance of biomes, as well as the foundation of biodiversity, based on rather suddenly evolutionary novelty, which challenges the Darwinian gradualism.

The symbiogenic concept allows for an innovative and broader approach to evolution, given that symbiosis is a fundamental rule in the establishment and development of life on Earth and elsewhere (Carrapiço et al. 2007). It implies a new paradigm for the comprehension of chemical and biological evolution. This change can be explained by a synergistic integrated cooperation between organisms, in which symbiosis acts, not as an exception, but rather as a rule in nature. We believe that competition and cooperation can co-exist in the same scenario of evolution, and probably take place in discontinuous bursts of activity, depending on the internal and external conditions that drive evolution. It means that the same population can evolve using competitive and/or cooperative processes during the time and space of a hypercycle evolutive scenario (Carrapiço et al. 2007). Thus, a series of synergistic and cooperative effects produced a wide source of creativity and functional advantages that pushed the emergence of complex and functionally integrated biological systems through the evolution of self-organization, self-catalysis and higher complexity.

These principles can be applied to the life on Earth and beyond, following the idea that the general mechanisms governing the evolutionary principles must be universal, surpassing the terrestrial dimension and projecting themselves in the universe. The biological organisms resulting from this process acquire characteristics inherent to the ecosystems where they are established and develop.

References

Bouchard, F. 2007. What is a symbiotic superorganism and how do you measure its fitness? A*bstracts of the ISHPSSB meeting*, University of Exeter, Exeter, UK, July 25–29, 2007, 45.

Carrapiço, F. 2010. How symbiogenic is evolution? *Theory in Biosciences* 129: 135–139.

Carrapiço, F., L. Pereira, and T. Rodrigues. 2007. Contribution to a symbiogenic approach in astrobiology. *Proceedings of SPIE* 6694: 669406-1–669406-10.

Chapman, M.J., and L. Margulis. 1998. Morphogenesis by symbiogenesis. *International Microbiology* 1: 319–326.

De Bary, A. 1878. Ueber Symbiose. – Tageblatt 51. Versamml. Deutscher Naturforscher u. Aerzte. *Cassel* 1878: 121–126.

Dubos, R., and A. Kessler. 1963. Integrative and disintegrative factors in symbiotic associations. In *Proceedings of the thirteenth symposium of the society for general microbiology*, eds. P.S. Nutman and B. Mosse, 1–11, London: University Press.

Lederberg, J. 1952. Cell genetics and hereditary symbiosis. *Physiological Reviews* 32: 403–430.

Margulis, L., and R. Fester. 1991. *Symbiosis as a source of evolutionary innovation*, Speciation and morphogenesis. Cambridge: The MIT Press.

Mayr, E. 1982. *The growth of biological thought. Diversity, evolution and inheritance*. Cambridge: Harvard University Press.

Mereschkowsky, C. 1905. Uber Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biologisches Centralblatt* 25: 593–604 (addendum in 25: 689–691).

Mereschkowsky, C. 1909. The theory of two plasms as foundation of symbiogenesis. A new doctrine on the origins of organisms. In *Proceedings studies of the imperial Kazan University*, vol. 12, 1–102, Publ. Office Imp. Univ, Kazan, Russia.

Sagan, L. 1967. On the origin of mitosing cells. *Journal of Theoretical Biology* 14: 225–274.

Sapp, J. 2003. *Genesis: The evolution of biology*. New York: Oxford University Press.

Sapp, J. 2004. The dynamics of symbiosis: An historical overview. *Canadian Journal of Botany* 82: 1046–1056.

Speidel, M. 2000. The parasitic host: Symbiosis contra neo-Darwinism. *Pli* 9: 119–138.

Chapter 8
Plant Neurobiology: Lessons for the Unity of Science

Paco Calvo Garzón

Abstract I propose to study the integration of contemporary scientific knowledge in cognitive neuroscience and plant neurobiology in order to assess the Unity of Science hypothesis. Oppenheim and Putnam (Unity of science as a working hypothesis. In: Feigl H, Maxwell G, Scriven M (eds) Minnesota studies in the philosophy of science. University of Minnesota Press, Minneapolis, pp 3–36, 1958. Reprinted in Boyd R, Gasper P, Trout JD (1991) The Philosophy of Science) considered the sort of mereological support that the Unity of Science hypothesis may receive from the principles of ontogenesis and evolution. I shall argue that a mechanistic understanding of eukaryote communication via the propagation of action potentials shows that the principle of ontogenesis does not support the hypothesis. Although the safest bet in my view is to press on a particular form of indirect evidence that the principle of evolution provides, I shall conclude that at present the Unity of Science remains an open empirical working hypothesis.

Keywords Evolution • Mechanistic explanation • Ontogenesis • Plant neurobiology

8.1 Introduction

According to Oppenheim and Putnam (1958, O&P), levels of organization in nature go from sociology and psychology all the way down to chemistry and the physics of fundamental particles. In particular, O&P distinguish six levels of organization: (6) Social groups; (5) (multicellular) living things; (4) cells; (3) molecules; (2) atoms;

P.C. Garzón (✉)
Departamento de Filosofía, Edificio Luis Vives, Universidad de Murcia,
Campus de Espinardo, E-30100 Murcia, Spain
e-mail: fjcalvo@um.es

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology, and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_8,

and finally (1) elementary particles (*ibid.*, p. 409). In their model, all five levels above the one represented by fundamental physics, are microreductive. That is, they are subject to reduction, and reduction proceeds downwards. In this way, any level in the hierarchy represents a potential microreduction of the next higher level. The Unity of Science, in its strongest form, represents the ideal that a genuine unification of scientific disciplines can be obtained by means of partial microreductions that will accumulate, eventually giving rise in the future to a basic understanding of all phenomena at a single bottom level of discourse.

O&P provide both direct as well as indirect factual support for the Unity of Science. On the positive side, they review, for instance, confirmatory evidence for the microreduction of level (5) to level (4). O&P have in mind the microreduction of the laws that govern animal behaviour to the microlaws that obtain at the cellular level. On the indirect side, they consider the sort of mereological support that *ontogenesis*, *evolution* and *synthesis* can lend to their working hypothesis. For present purposes, I shall ignore paradigmatic examples of synthesis, such as the pharmacological synthesis of molecules out of the chemical manipulation of atoms, and focus instead on the first two forms of indirect support (ontogenesis and evolution).

As an example of ontogenesis, O&P contemplate the case of amoebae like the slime mold. At one point, these protozoa get clustered, forming multicellular organisms of a considerable degree of sophistication, as illustrated by the fact that they show signs of a directional response to light sources (Bonner 1952). In O&P's view, the example of the slime mold partially vindicates their hierarchy of microreductions insofar as the genesis of the multicellular organisms is the next lower level, in this case, the level of unicellular species. Likewise, as we move to other levels in the microreductive hierarchy we find that the same ontogenetic analysis can be invoked. Atoms, for instance, come out of clusters of elementary particles, and the same goes for any two adjacent levels in O&P's hierarchy.

On the other hand, with regard to evolution, O&P claim:

> Evolution … is an over-all phenomenon involving all levels, from 1 to 6. [Whenever] it can be shown that things of a given level existed before things of the next higher level came into existence … some degree of indirect support is provided to the particular special case of our working hypothesis that concerns those two levels … Wherever one draws the line, non-living molecules preceded primordial living substance, and the latter evolved gradually into highly organized living units, the unicellular ancestors of all living things. (*ibid.*, 418–9)

In O&P's view, their working hypothesis thus receives a form of indirect support from mereological considerations that relate to ontogenesis and evolution. On the one hand, at one point in ontogenesis, a systematic aggregation at one specific level did not exist, whereas its constitutive parts, at the next level down the hierarchy, did exist. On the other hand, life evolved in complexity in such a way that at certain times in the history of the universe, systems at a given level of aggregation existed, but not at higher levels. Ontogenesis and evolution, in short, are hypotheses as to the genesis of increasingly complex levels of aggregation, both in the context of the individual and in the context of the life of the universe. The question is whether granting ontogenesis and evolution to be well-confirmed theoretical hypotheses, we thereby have a direct route to scientific unification via microreduction at the lowest level of physical aggregation in O&P's ontic hierarchy.

Recent research (Pruitt et al. 2003; Trewavas 2005) shows that the behaviour of eukaryotic organisms connects with the molecular level in a uniform manner. Insofar as plants and animals belong to the eukaryotic kingdom, a domain whose phylogeny has been hypothetically tracked down to a common universal ancestor in the domain of bacteria (Campbell 1996), we have a reason, I propose, to study the integration of contemporary scientific knowledge in cognitive neuroscience (Gazzaniga et al. 2002) and plant *neurobiology* (see below) in order to assess the Unity of Science hypothesis.

Evolutionarily speaking, plants and animals diverged in obvious ways. Animals, insofar as they are heterotrophic organisms that require organic foodstuff to survive, exploit a number of mobility-related competencies in order to navigate in complex and contingent environments (Neumann 2006). Plants, on the other hand, do not require contractile muscles for fast responses to environmental contingencies. Insofar as plants are autotrophic organisms, they operate in slower time scales since inorganic substrates can be synthesized while remaining stationary. Similarities, nonetheless, also exist. Although there is no scientific evidence of the existence of brains, nerves and synapses in plants (Alpi et al. 2007), auxin transport in plant cells can be functionally equated with neuronal networks in animals (see Brenner et al. 2007; Trewavas 2007). In fact, as we shall see below, the evolution in vascular plants of roots and tissues, which allow for the circulation of substrates throughout the plant system, allows us to interpret their adaptive responses in neurobiological terms (see Calvo Garzón 2007; Calvo Garzón and Keijzer 2011).

In my view, granting this embracing picture, the safest bet for the sympathizer of O&P is to put in perspective eukaryotes with regard to their shared unicellular ancestors, pressing thus on further indirect factual evidence from evolution. Once we look at the shared cellular and molecular mechanisms of some eukaryote life forms, we may have a reason to unify the knowledge obtained along the spectrum. However, the fact that such a microreductive strategy has already delivered considerable goods and will certainly continue in the future does not necessarily back the Unity of Science. It might furnish inductive evidence in favour of the *Unity of Science* as "a pervasive *trend* within science" (*ibid.*, p. 406), but it is still compatible with a reductionist strategy orthogonal to O&P's stronger reading ("an ideal state of science", *ibid.*, p. 406).

In this chapter I shall argue that a mechanistic understanding (Bechtel and Richardson 1993) of eukaryote communication via the propagation of action potentials shows that ontogenesis cannot lend any sort of inductive support to the Unity of Science. Furthermore, although the safest bet for a sympathizer of O&P is to press on the indirect evidence that evolution provides, I shall conclude that their working hypothesis remains an open empirical question.

8.2 Plant Neurobiology: A Mechanistic Approach to Ontogenesis

Plant neurobiology (Baluška et al. 2006; Brenner et al. 2006) is an ever-growing discipline whose aim is the scientific understanding of integrated plant behaviour.[1] As an academic discipline it has emerged only in recent years as a result of the incorporation of new knowledge from well established areas of research such as plant electrophysiology, cell biology, molecular biology, and ecology. As in the case of cognitive neuroscience (Gazzaniga et al. 2002) in relation to, say, neurophysiology, plant neurobiology differs from plant physiology (Taiz and Zeiger 2002) insofar as the emphasis is laid in the interdisciplinary effort whose ultimate target is the study of the complex patterns of behaviour of plants *qua* information-processing systems. Plant neurobiology interprets plants as information-processing networks with individual cells as computational building blocks. The following quotation from a recent survey of the state of the art in the literature clearly shows that the neurocomputational features of plants are meant to be taken literally:

> Each root apex is proposed to harbour brain-like units of the nervous system of plants. The number of root apices in the plant body is high, and all "brain units" are interconnected via vascular strands (plant neurons) with their polarly-transported auxin (plant neurotransmitter), to form a serial (parallel) neuronal system of plants. (Baluška et al. 2006, 28).

Plant neurobiology has identified many components of mammal nervous mechanisms (Baluška et al. 2004), and a lot is known about the molecular constituency of cellular level processes in plants (Baluška et al. 2006). Plant cells need to synchronize their inner doings with environmental regularities. This is achieved via communication channels at the molecular level. Calcium waves open up the possibility to integrate and compute all incoming data (Trewavas 2002). Mechanistic explanations (Bechtel and Richardson 1993), furthermore, can be used to identify components that get organized spatially and temporally for signalling purposes.

With regard to mechanisms, Craver and Bechtel (2006) (see also Craver 2009) distinguish between phenomenal, componential, causal, and organizational aspects, and illustrate the approach with the identification of ion mechanisms as the components of action potentials (APs) in animal neuronal cells. APs consist of electrical discharges that propagate throughout cells' membranes. The generation of action potentials is the *phenomenon* that the ion mechanism explains in terms of the interactions among its *components*. And crucially, these components *causally* interact in an organized manner, in such a way that a number of *spatiotemporal* constraints are critical to the manifestation of the phenomenon under study; in our case, the generation of APs.

Ion channels implement excitability in cellular tissues in animals and plants, and a mechanistic framework will allow us to compare ion mechanisms as the components of APs' signalling across eukaryotes. In plants, communication is

---

[1] For an introduction to plant neurobiology, see http://www.plantbehavior.org/neuro.html

achieved via APs that propagate by means of phloem cells.[2] In both animals and plants, ion channels are the key component in the generation of APs. Differing ion concentrations at both sides of the cell's membrane cause a potential difference between the inner and the outer part of the membrane. When the membrane is in its resting state, different ion distributions obtain at the outer and inner parts of the membrane. APs consist of three clearly differentiated phases: First, (a) a depolarization obtains, with an increase of electric potential; (b) the potential then falls quickly below the original resting level (repolarization); and finally, (c) a hyperpolarization takes place, a period in which the cell remains less excitable.

In order to understand mechanistically the three-fold process that gives rise to the propagation of APs, it is essential that we identify the components of the generating mechanism, and the way they get arranged both spatially and temporally. In animals, the components involved in APs are sodium and potassium ions, mainly, and a semipermeable cellular membrane with voltage-selective ion channels that permit the flux of ions inwards and outwards. APs only generate out of particular causal connections among these components. Critically, specific spatiotemporal constraints explain the three phases involved. These constraints relate to the particular arrangement of ion channels across the membrane, and more importantly, to the temporal ordering of events that take place in relation to the activation of specific channels (see Craver and Bechtel 2006, for the details). These transient changes in membrane conductance due to the specific organization of the components in the mechanism are the cause of the typical depolarization-repolarization-hyperpolarization shape of APs' waveforms (Hodgkin et al. 1949).

In plants, on the other hand, AP long-distance signalling is critical for the regulation of respiration, photosynthesis and gene expression, among other things. APs involve ion exchanges of calcium, and chloride and potassium. The ion components of APs are well known from studies of giant *Characean* cells. In particular, at the cellular resting state, there is a uniform distribution of Ca^{2+} and Cl^{-}, which are kept far from their electrochemical equilibrium at the outer and inner parts of the membrane, respectively. The AP ion mechanism involves an increase in the concentration of Ca^{2+} in the cytoplasm what in turn causes an efflux of Cl^{-} ions that leak out down their potential gradient. This triggers a depolarization phase at the membrane (Trebacz et al. 2006). Repolarization takes place with the opening of voltage-gated potassium channels that permit K^{+} to leak out as well. The hyperpolarization phase is induced by the increase of K^{+} influx (as happens to be the case in animal nerve cells), and the resting potential gets restored via an electrogenic proton pump.

This brief review of the ion mechanisms involved in the generation of APs in animals and plants will allow us to consider O&P's Unity of Science hypothesis,

---

[2] The crucial difference between APs mechanistic components in animals and plants is that the electric profile of APs in the former is implemented, mainly, via potassium and sodium channels, whereas in the latter case, potassium, chloride, and calcium channels are primarily involved. For a classic review of plant APs, see Pickard (1973).

first of all, in terms of the alleged support it receives from ontogenesis, and, in the next section, in terms of the alleged support it receives from evolution.

In motivating their particular hierarchy and microreductivist approach, O&P introduce a number of adequacy conditions, more or less contentious, such as the assertion that a number of levels must exist, or that levels cannot be infinite in number. For present purposes, I shall concentrate on conditions (iii) and (iv) which are more controversial. According to condition (iii), "There must be a unique lowest level" (*ibid.*, p. 409), and according to condition (iv), "Any thing of any level except the lowest must possess a decomposition into things belonging to the next lower level" (*ibid.*, p. 409).

It is not clear, however, that ontogenesis can inductively support the Unity of Science, or at least it would require further motivation. The explanatory effort that ontogenesis points towards is compatible with a reductionist understanding of the enterprise of scientific research that may be alien to O&P's initial formulation. The motivation for this alternative picture has to do with the mechanistic conception of scientific explanation favoured here, as illustrated by the generation of APs in animals and plants.

In particular, once we adopt a mechanistic conception of scientific explanation, conditions (iii) and (iv) are not met. As Craver and Bechtel (2006) point out, the identification of components within a mechanism only makes sense in the context of the phenomenon we aim to explain. Aggregated components do not thereby become components-of-the-mechanisms just by being building blocks in terms of ontogenesis. Rather, parts must be causally involved in the spatiotemporal sequence that gives rise to the phenomenon in question. If the specific sequence in terms of ion channel conductance, for example, that takes place in the cellular membrane was irrelevant to the waveform obtained, it would fail to be a component of the mechanism that explains the generation of the AP.

The moral is that the hierarchical relations that emerge, as components in mechanisms are identified in relation to the phenomenon they produce, need not coincide with O&P's ontic hierarchies. The latter sort of hierarchies is posited, as the argument from ontogenesis stresses, because of the very fact that things came out of things at lower levels. This ontic-based relation however works irrespective of explanatoy needs. It is a byproduct of the way levels of aggregation in nature happened to arrange. Crucially, the way a mechanism decomposes depends upon the phenomenon in need of mechanistic explanation. This way, if in terms of ontogenesis, something is formed out of smaller parts, those parts will not play any role in the mechanistic hierarchy since they are not proper components in the mechanism that explains the phenomenon.

But, can the sympathiser of the Unity of Science deliver the goods by pressing from the side of evolution? At first sight such an approach might be subject to the very same criticism raised against ontogenesis. After all, the fact that an individual can be decomposed into components at different levels as a function of the phenomenon in need of explanation, has to do with how mechanistic explanations operate, and not with the fact that decomposition proceeds at the level of ontogenesis. In this way, a parallel argument can easily be phrased for evolution: Although life evolved in complexity in such a way that at certain times in the history of the universe, systems at a given level of aggregation existed, but not at higher levels, there may be more than one way to decompose a system at one point in evolution. As in the case of individuals under the ontogenesis lens, it simply depends upon our explanatory needs. Put bluntly, mereological parts do not necessarily correspond with the posits of successful microreducing theories.

In my view, however, a different form of indirect support from evolution that, to the best of my knowledge, has not been echoed in the literature may be provided by the fact that different life forms share lower-level mechanisms. Rather than the result of a cosmic fluke, evolution is a plausible theoretical hypothesis that explains how very different life forms, such as plants and animals, or for that matter, even fungi, happen to share such lower-level mechanisms.[3] With regard to condition (iii), O&P claim that "success at transforming all the *potential* microreductions connecting these branches into *actual* microreductions must, *ipso facto*, mean reduction to a single branch" (*ibid.*, p. 409). Although in principle, I agree, this alleged "reduction to a single branch" is questionable, the fact that lower-level mechanisms have been identified along philogenia, may furnish evidence of a different sort, as I shall argue in what follows.

8.3 Evolution and Anticipatory Behaviour

Although ontogenesis by itself does not seem to lend support to the Unity of Science, there is a further argument, this time alluding to evolutionary considerations, that may back O&P's working hypothesis somewhat more indirectly. A well-confirmed hypothesis that may lend credibility to the Unity of Science is that all eukaryotic life forms evolved out of one single ancestral cell type shared by all of them. One reason to believe in this hypothesis is that animals and plants share a vast number of basic molecules that relate to cell division, among other life-supporting doings, such as respiration. As we saw earlier, at the level of the ion mechanisms, a calcium ion influx, and a potassium and chloride ion efflux are produced. In this way, we may say that whereas in animal nerve cells, depolarization involves an increased influx of $Na+$ ions, APs in plants involve an influx of Ca^{2+} and an efflux of Cl^- ions (Stahlberg 2006). Crucially, independently of the ion channels involved in depolarization, most of the fundamental properties of APs in animals are shared by plant APs.

The resemblance between animal and plant excitability is straightforward. After all, cellular excitation is based upon chemical protoplasmatic activity, prior to the eukaryotic differentiation between plant and animal tissues took place. APs are the result of the depolarization that takes place in membrane potentials of nerve

---

[3] Ion channels implement excitability in cellular tissues in animals, fungi, and plants. As a matter of fact, they can be traced back to bacteria in philogenia (Hille 2001).

cells and phloem cells in animals and plants, respectively. The key issue is that eukaryotic bioelectrochemical signalling is present everywhere. Animal neurons have an excitable membrane that permits the propagation of action potentials, and, likewise, phloem cells have a membrane that permits the production of action potentials in plants.

APs in animals and plants share some core properties with voltage-gated ion channels being critical in the induction of APs.[4] Independently of the implementation details of the ion mechanism involved, long-distance transmission of signals in plants functions via phloem[5] similarly as it does axonically in animal neural networks. With regard to electrophysiological properties, APs behave in an all-or-none fashion, giving rise to constant, full amplitudes (Brenner et al. 2006); they propagate without decrement; there are refractory periods in which excitability decreases and the propagation of signals is momentarily prevented; and of course, all three phases in the characteristic waveform are contemplated. That is, transient membrane depolarizations are followed by repolarization at a second stage, after which hyperpolarization obtains. If we are to find a basic process shared by all living organisms that is the conduction of electrochemical signals as the result of excitation.[6]

However, the sort of support I wish to evaluate from evolution does not stem from the fact that higher-level intelligence seems to derive in an accumulative manner out of the evolutionary adaptations of lower-level mechanisms. That is O&P's reading, but such a reading is to some extent subject to the criticism previously raised against ontogenesis, as we saw earlier.

Insofar as eukaryotes' information-processing capacities depend upon the details of their neurocomputational architecture, we can attempt at explaining the behavioural output of plants and animals by identifying components and the way they get organized. What really matters, I contend, in assessing the parallelism between specific patterns of behaviour in plants and animals is (i) whether there is some type of competency that can be said to be shared by them, and (ii) whether that competency gets implemented in similar ways at an appropriate level of aggregation. However, if we are to find a competency that permits the comparison among eukaryotic life forms, it would better be of a sufficient degree of complexity so that it remains explanatorily adequate in the context of animal behaviour.

---

[4] For a comparison of APs in animal and "plant neurons" in *Nitella* and the giant axon of squids, see Cole and Curtis (1938, 1939). For a brief but interesting historical overview of the concept of nervous systems in plants, see Stahlberg (2006).

[5] APs propagate from stem to root and viceversa via the phloem sieve-tube system.

[6] Obviously, were we to consider the modelling and quantification of specific APs in animals and plants, a number of differences would be found. Resting potentials in *Chara* cells, for example, are more hyperpolarized than their counterpart in animal cells. Also, depolarization is induced by anions, rather than by cations as is the case in the triggering of depolarization in animal APs (see Kikuyama 2001). In addition, APs amplitudes and refractory periods are significantly larger in *Chara*, taking up to several seconds, whereas they last msecs in animal nerve cells.

The question then is: What sort of competency can lend support by unifying our explanatory understanding of animal and plant based overt behaviour? Plant neurobiologists consider plants' ability to integrate exogenous and endogenous information channels in an attempt to phenotypically adapt to environmental contingencies a sophisticated form of competency that can be paralleled to animals' predictive behaviour. I propose to consider de-coupled, off-line modelling tasks, since they are usually taken to be what distinguishes sophisticated forms of behaviour from merely reactive (online) routines.[7] In my view, the real reason to move from ontogenesis to evolution is that different organisms seem to share off-line competencies that get explained mechanistically in the same way. Although out of one starting point of eukaryotic unity, different "solutions" emerged,[8] we can explore whether functionally similar off-line competencies in both animals and plants can lend support to the Unity of Science.

In this way, illustrations of plant behaviour functionally similar to O&P's example of the slime mold will certainly not help. Although the degree of sophistication as a result of multicellular clustering in protozoa is considerable, a plant counterpart of the slime mold would not constitute a good example, since its ability to respond directionally to light sources only constitutes a reactive online competency.

Fortunately, as research in plant neurobiology shows, plants are not simple systems that passively build up photosynthate. Rather, plants can sense volume, discriminate own from alien root structures, and allow for phenotypic root reordering as a function of competition for nutrients, to list but a few examples. Many more sophisticated plant competencies could be listed (see Trewavas 2005, 2006),[9] but for the purposes of identifying a form of off-line anticipatory behaviour in plants, I shall focus on plant tropisms. In the remaining of this section, I shall elaborate on tropistic responses in order to show how a mechanistic approach to scientific explanation allows us to confront the way neural networks get implemented in animals and plants.

Generally speaking, tropisms involve a directional change such as growth or movement in response to a given environmental stimulation (Stanton and Galen 1993). Plant tropisms (Koller 1986) however can vary substantially as a function

---

[7] In fact, it has been suggested many times that the *sine qua non* of representation-based competency is off-line adaptive behaviour (see, for example, Clark 1997). For a critical appraisal of Clark's suggestion in the context of representational theories of cognition, see Calvo Garzón (2008).

[8] For plant life based on photosynthesis there was no need to evolve locomotion. In this way, whereas animals found a heterotrophic way out of their energy-consumption needs (hunting, etc.), plants found an autotrophic solution (motionless organic synthesis) (Trewavas 2002).

[9] Carnovorous *D. muscipula* and *A. vesiculosa*, for example, would furnish us with a primitive form of plant memory subject to a mechanistic interpretation in the context of off-line adaptive behaviour. In the case of *D. muscipula*, an AP is generated whenever an upper trap hair is bent. Crucially, one single stimulation of the hair does not trigger the closing of the trap. For the trap to close it is necessary a second AP that takes place only when another hair is bent within 40 s after the first AP has been generated (see Baluška et al. 2006, for the details). We may then interpret the second AP as a primitive form of plant memory.

of the type of stimulus that the plant is responsive to, and the part of the plant that responds to the stimulation. Thus, we may talk of gravitropisms, when roots react to gravity; thigmotropism, in response to physical contact; or directional responses to moisture (hydrotropism), and temperature (thermotropism), for example. Complex, although still online, tropistic reactions are frequent. Roots that manifest gravitropism stop developing downwards as they encounter a physical obstacle, and grow horizontally instead. However, they are able to assess online the state of affairs, and periodically they try to move downwards, and remain horizontal is unable to respond gravitropically (see Massa and Gilroy 2003).

In addition, plants exhibit more sophisticated forms of behaviour, being able to assess current data that can suppose an advantage at a later stage. Roots, for instance, exhibit patterns of growth that depend upon future acquisition of minerals and water.[10] Plants, in short, can model environmental regularities in order to predict the future, and selectively change phenotypically so as to achieve distant goals towards global fitness. Such a modelling behaviour is certainly a stepping stone that distances plant off-line computational capabilities from merely reactive (online) life forms. This constitutes the basis of a predictive form of plant behaviour that can be assessed in relation to the Unity of Science. In particular I wish to focus on leaf heliotropism; a specific form of light-related tropistic behaviour.

In relation to light-related tropisms, it is important to distinguish phototropisms, which may represent directed responses to a static light source, from heliotropism, a more complex response that involves a correlated response to changes in sunlight orientation as the day changes from sunrise to sunset. We may also distinguish between flower heliotropism and leaf heliotropism. The reason to draw this distinction is that in the case of flower heliotropism, no "memory" mechanisms seem to be required for flowers to keep track of the position of the sun. Unless flowers are exposed to light in the morning they will fail to reorient to sunrise, remaining in a random orientation throughout the night.

Off-line nocturnal reorientation by plant leaves represents a qualitative change with regard to more reactive behaviour. Leaf laminas of *Lavatera cretica* can, not only anticipate the direction of the sunrise, but also allow for this anticipatory behaviour to be retained for a number of days in the absence of solar-tracking (Schwartz and Koller 1986). That is, the laminas reorient during the night and keep facing the direction of the sunrise even after a few days without tracking the sun, and without sensing the position of sunset. An experiment reported by Schwartz and Koller (1986) clearly shows that this is a complex off-line response.

In a series of experiments, three groups of plants were taken at sunset to three different cabinets. One was kept in darkness; another had illumination in day-light

---

[10] Functionally similar results obtain in the case of shoots. Other typical examples include making decisions about the number of flowers to produce 1 year in advance of the flowering season, or branching-related decisions made with in some cases years of anticipation. Some plants predict potential future shades out of reflected far-red/red light. Also, some trees synchronize their metabolic activity with non-drought periods after exposition to long epochs of rain drought (see Trewavas 2005, and references therein).

hours, and another one was kept illuminated vertically throughout the experiment. The first day, plants in cabinets 1 and 2 (darkness and day-light hours conditions) could anticipate sunrise and leaves were oriented towards that direction. Plants in cabinet 3, on the other hand, were horizontal, with laminas facing upwards, as they had been kept under constant vertical artificial illumination. The nocturnal reorientation behaviour exhibited by plants in cabinets 1 and 2 lasted for as long as 3 days under the same experimental conditions, that is, in the absence of day-time solar-tracking.[11] In their study, the explanation of nocturnal reorientation cannot be the sunrise of the day before, since plants were prevented from tracking the sun for 3–4 days.

Crucially, the explanation of nocturnal reorientation involves the internal modelling of environmental rhythms. Circadian clocks allow for time-estimation by synchronizing endogenously generated activity with exogenous cyclic periods such as day-night planetary patterns. Circadian clocks can mimic biological rhythms on a 24-h cycle, and this explains nocturnal reorientation in plants for up to 4 days in the absence of sunrise stimulation.

Plant genetics points towards underlying shared molecular components that mechanistically explain day-length estimations and the operation of light receptors (see Pruitt et al. 2003, and references therein). In the case of time-estimation, recent research in genomics has unearthed the molecular mechanisms of plant overt behaviour, with the striking result that both plants and animals draw on the very same molecular networks (Cashmore 2003) in their adaptive exploitation of circadian clocks. One single level, thus, explains the origins of the anticipatory behaviour as exemplified by circadian clocks. This research points towards a type of (molecular) *unification* that *may* be compatible with O&P.

As far as O&P's evolutionary argument is concerned, time-estimation has proven critical to all sorts of eukaryotic organisms, since synchronization with exogenous cycles allows organisms to anticipate regular or quasi-regular ecological changes that may have a selective value. What unites animals and plants in evolutionary terms is the need to exploit an internal memory that allows organisms to plastically change their behaviour in order to optimize fitness. That points towards a form of memory and learning shared by eukaryotes. Short-term memory in *Aplysia*, for instance, permits this marine snail to avoid danger in much the same way that drought is avoided by plants (see Trewavas 2003). Once we stick to the cellular and molecular level, things do not look that different across the eukaryotic kingdom. As Pruitt et al. (2003) show, numerous epigenetic phenomena do have in common a number of cellular underlying mechanisms.

An inference to the best explanation would point towards a rather premature starting point in the configuration of such mechanisms. Time-estimation gives rise to a primitive form of anticipatory behaviour that has proven critical in

---

[11] Since plants in cabinet 2 had only been illuminated vertically, the re-orientation exhibited could only be due to the sunrising information being stored in advance.

phylogenesis.[12] Nervous systems in animals then diverged at some point in the evolutionary trajectory, due to different pressures and needs from those of sessile plants. Nevertheless, the molecular level in O&P's hierarchy furnishes us with an embracing picture of eukaryotic anticipatory capacities; a framework that may allow us to place amoebae, plants, and animals (human and non-human) along a continuum. In fact, once we look at the shared cellular and molecular mechanisms of these life forms, we have a reason to unify the knowledge obtained along the spectrum. The crucial point is that molecules that evolutionarily speaking preceded sophisticated living forms are part of the equation that mechanistically explains a competency that is shared by those live forms.

Plant phenotypic changes are to lower level molecular mechanisms what stable patterns of animal overt behaviour are to endogenous molecular activity. Beyond ontic hierarchies, such a level of explanation serves to unify certain off-line adaptive forms of behaviour. However, such an embracing picture does not necessarily lend support to the Unity of Science hypothesis. The fact that molecules were present before multicellular organisms came into existence is compatible with a mechanistic reading of the components involved that might not back up the micro-reductive link between the molecular level and the level of (multicellular) living things in O&P's hierarchy.

Certainly, we can actually try to get a fix on the underlying mechanisms that trigger the patterns of behaviour, but before we have a microreductive model up-and-running we must set up constraints on which diverse levels of analysis are appropriate to study the phenomenon in question. That is, we must come to a decision as to whether for a neuroscientific understanding of a cognitive task, for example, the biochemistry of neurotransmitters needs to be honoured or whether modelling the firing rates of populations of, say, motor neurons suffices.[13] And the same goes for plant behaviour. How do we know in general that any level in O&P's hierarchy is essential to a full-fledged explanation of behavioural output? The litmus test that allows us to settle at a given level of aggregation is the existence of linkages between that level in question and the overt behaviour we aim to explain, be it nocturnal plant reorientation or long-term potentiation in animals.

Although in principle, I agree, the alleged "reduction to a single branch" is questionable, the fact that lower-level mechanisms have been identified along philogenia,

---

[12] Circadian clocks are a wide-spread trick that evolution found in order to keep track of the rythms of nature in the absence of direct stimulation. In evolutionary terms, circadian clocks have emerged as a minimum of four times (see Dodd et al. 2005). It is then clear that organisms endowed with such an estimate of ecological cycles must have some sort of Darwinian advantage.

[13] Bickle (2003) has recently presented an illustration of a successful linkage between lower level molecular mechanisms and stable overt behaviour. He contends that an explanation of 'quantitative behavioral data at the level of biochemical pathways and intracellular molecular mechanisms' is already on offer. Specifically, he shows how particular behavioural data can be explained at the level of the molecular mechanisms that implement long-term potentiation (LTP). For a discussion of linkages between lower and higher levels in the context of dynamicism, see Calvo Garzón (2008).

may furnish evidence of a different sort. Unfortunately, although it is clear that APs relate to light-induced tropisms in general, a lot is still missing in order to understand the precise molecular basis of ion mechanisms, and the association between electric signalling and leaf heliotropism. As more knowledge is gathered in areas such as genomics, plant electrophysiology, and bioinformatics, a better understanding of the role played by sieve tubes and plasmodesmata in the propagation of APs, and more generally of how molecular and cellular networks relate to plants' overt behaviour in order to fully appraise how, for example, long-distance signalling regulates their adaptive responses to environmental contingencies, will be obtained. Put bluntly, the aforementioned linkages are still not known in sufficient detail. The same can be said with respect to other existing analogies between animals and plants, such as plant "synapses" (the "vesicle-operated intercellular clefts in axial root tissues", Stahlberg 2006)

8.4 Conclusion

In this chapter I have considered the integration of contemporary scientific knowledge in cognitive neuroscience and plant neurobiology in order to assess the support that the Unity of Science hypothesis may receive from the principles of (i) ontogenesis and (ii) evolution.

(i) With regard to ontogenesis, the hierarchical relations that emerge, as components in mechanisms are identified in relation to the phenomenon they produce, need not coincide with O&P's ontic hierarchies. The adequacy conditions that O&P's line of argumentation relies upon are thus not met. In particular, it is not clear that decomposition characterizes the relation between any two contiguous levels in O&P's hierarchy. Aggregated components do not become components-of-the-mechanisms just by being building blocks in terms of ontogenesis. Ontogenesis, I concluded, does not lend inductive support to the Unity of Science.

(ii) In relation to the principle of evolution, the sort of support that I have considered does not stem from the fact that higher-level intelligence derives in an accumulative manner out of the evolutionary adaptations of lower-level mechanisms. That is O&P's reading, but such a reading is subject to the same criticism being raised against ontogenesis.

On the other hand, as I argued, although ontogenesis and evolution by themselves do not seem to lend support to the Unity of Science, the principle of evolution may back O&P's working hypothesis somewhat more indirectly. In particular, although out of one starting point of eukaryotic unity, different "solutions" emerged in terms of adaptability, we can explore whether functionally similar off-line competencies are shared by animals and plants. The crucial point is that molecules that evolutionarily speaking preceded sophisticated living forms are part of the equation that mechanistically explains such competencies. I proposed to consider de-coupled, off-line modelling tasks, since they are usually taken to be what distinguishes sophisticated forms of behaviour from merely reactive (online) routines. The real reason to move from ontogenesis to evolution then is that different organisms seem to share off-line competencies that get explained mechanistically in the same way.

(Un)fortunately, evolution is not a definite argument for scientific unification via accumulative microreductions to a single bottom level, but rather an accomplished partial microreduction that points towards shared mechanisms. If shared mechanisms happen to be present at lower levels (atoms, etc., etc.) and linkages to overt behaviour are equally obtained, then further evidence for microreduction will be available. In this way, intertheoretic reduction is compatible with ontic basement levels other than the bottom one in their 6-level hierarchy. Whereas O&P see the process of scientific unification as one where partial microreductions drive scientific knowledge ultimately to the level of particles, I regard particular episodes of microreductions compatible with a number of bottom levels. Which levels these happen to be the case will need to be decided on a one-to-one basis.

The linkage between that level in question and the overt behaviour we aim to explain will allow us to settle at a given level of aggregation. At this point, more research is needed in order to assess whether the evidence provided inductively supports the *Unity of Science* only as "a pervasive *trend* within science" (*ibid.*, p. 406), or, more strongly, as "an ideal state of science", (*ibid.*, p. 406). The issue, at present, remains an open empirical question.

Acknowledgements I would like to thank Paul Humphreys, John Symons and Anthony Trewavas for their comments on a previous version of this manuscript. Preparation of this manuscript was supported by DGICYT Projects HUM2006-11603-C02-01 (Spanish Ministry of Science and Education and Feder Funds) and FFI2009-13416-C02-01 (Spanish Ministry of Science and Innovation), and by Fundación Séneca-Agencia de Ciencia y Tecnología de la Región de Murcia, through project 11944PHCS 09.

References

Alpi, A., et al. 2007. Plant neurobiology: No brain, no gain? *Trends in Plant Science* 12: 135–136.

Baluška, F., S. Mancuso, D. Volkmann, and P. Barlow. 2004. Root apices as plant command centres: The unique 'brain-like' status of the root apex transition zone. *Biologia (Bratisl.)* 59: 9–17.

Baluška, F., A. Hlavacka, S. Mancuso, and P. Barlow. 2006. Neurobiological view of plants and their body plan. In *Communication in plants: Neuronal aspects of plant life*, ed. F. Baluška et al. Berlin/Heidelberg: Springer.

Bechtel, W., and R.C. Richardson. 1993. *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton: Princeton University Press.

Bickle, J. 2003. *Philosophy and neuroscience. A ruthlessly reductive approach*. Dordrecht: Kluwer.

Bonner, J.T. 1952. *Morphogenesis*. Princeton: Princeton University Press.

Brenner, E.D., R. Stahlberg, S. Mancuso, J. Vivanco, F. Baluška, and E. van Volkenburgh. 2006. Plant neurobiology: An integrated view of plant signalling. *Trends in Plant Science* 11: 413–419.

Brenner, E.D., R. Stahlberg, S. Mancuso, F. Baluška, and E. Van Volkenburgh. 2007. Plant neurobiology: The gain is more than the name. *Trends in Plant Science* 12: 135–136.

Calvo Garzón, P. 2007. The quest for cognition in plant neurobiology. *Plant Signaling & Behavior* 2: 1–4.

Calvo Garzón, P. 2008. Towards a general theory of antirepresentationalism. *The British Journal for the Philosophy of Science* 59: 259–292.

Calvo Garzón, P., and F. Keijzer. 2011. Plants: Adaptive behavior, root brains and minimal cognition. *Adaptive Behavior* 19: 155–171.

Campbell, N.A. 1996. *Biology*, 4th ed. Menlo Park: Benjamin/Cummings.

Cashmore, A.R. 2003. Cryptochromes: Enabling plants and animals to determine circadian time. *Cell* 114: 537–543.

Clark, A. 1997. *Being there: Putting brain, body and world together again*. Cambridge: MIT Press.

Cole, K.S., and H.J. Curtis. 1938. Electric impedance of Nitella during activity. *Journal of General Physiology* 22: 37–64.

Cole, K.S., and H.J. Curtis. 1939. Electric impedance of the squid giant axon during activity. *Journal of General Physiology* 22: 649–670.

Craver, C.F. 2009. Levels of mechanisms: A field guide to the hierarchical structure of the world. In *Companion to the philosophy of psychology*, ed. J. Symons and P. Calvo. New York: Routledge.

Craver, C.F., and W. Bechtel. 2006. Mechanism. In *Philosophy of science: An encyclopedia*, ed. S. Sarkar and J. Pfeifer, 469–478. New York: Routledge.

Dodd, A., N. Salathia, A. Hall, E. Kévei, R. Tóth, F. Nagy, J. Hibberd, A. Millar, and A. Webb. 2005. Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science* 309: 630–633.

Gazzaniga, M., R. Ivry, and G. Mangun. 2002. *Cognitive neuroscience: The biology of the mind*. London: Norton.

Hille, B. 2001. *Ion channels of excitable membranes*. New York: Sinauer.

Hodgkin, A.L., A.F. Huxley, and B. Katz. 1949. Ionic currents underlying activity in the giant axon of the squid. *Archives des Sciences Physiologiques* 3: 129–150.

Kikuyama, M. 2001. Role of Ca2+ in membrane excitation and cell motility in characean cells as a model system. *International Review of Cytology* 201: 85–114.

Koller, D. 1986. The control of leaf orientation by light. *Photochemistry and Photobiology* 6: 819–826.

Massa, G., and S. Gilroy. 2003. Touch modulates gravity sensing to regulate the growth of primary roots of Arabidopsis thaliana. *The Plant Journal* 33: 435–445.

Neumann, P.M. 2006. The role of root apices in shoot growth regulation: Support for neurobiology at the whole plant level? In *Communication in plants: Neuronal aspects of plant life*, ed. F. Baluška et al. Berlin/Heidelberg: Springer.

Oppenheim, Paul, and Hilary, Putnam. 1958. Unity of science as a working hypothesis. In *Minnesota studies in the philosophy of science*, ed. Herbert Feigl, Grover Maxwell, and Michael Scriven, 3–36. Minneapolis: University of Minnesota Press. Reprinted in Richard Boyd, Philip Gasper, and J.D. Trout *The philosophy of science*, 1991.

Pickard, B.G. 1973. Action potentials in higher plants. *The Botanical Review* 39: 172–201.

Pruitt, R., J. Bowman, and U. Grossniklaus. 2003. Plant genetics: A decade of integration. *Nature Genetics* 33: 294–304.

Schwartz, A., and D. Koller. 1986. Diurnal phototropism in solar tracking leaves of lavatera cretica. *Plant Physiology* 80: 778–781.

Stahlberg, E. 2006. Historical overview on plant neurobiology. *Plant Signaling & Behavior* 1(1): 6–8.

Stanton, M.L., and C. Galen. 1993. Blue light controls solar tracking by flowers of an alpine plant. *Plant, Cell & Environment* 16: 983–989.

Taiz, L., and E. Zeiger. 2002. *Plant physiology*, 3rd ed. New York: Sinauer.

Trebacz, K., H. Dziubinska, and E. Krol. 2006. Electrical signals in long-distance communication in plants. In *Communication in plants: Neuronal aspects of plant life*, ed. F. Baluška et al. Berlin/Heidelberg: Springer.

Trewavas, A. 2002. Mindless mastery. *Nature* 415: 841.

Trewavas, A. 2003. Aspects of plant intelligence. *Annals of Botany* 92: 1–20.
Trewavas, A. 2005. Green plants as intelligent organisms. *Trends in Plant Science* 10(9): 413–419.
Trewavas, A. 2006. The green plant as an intelligent organism. In *Communication in plants: Neuronal aspects of plant life*, ed. F. Baluška et al. Berlin/Heidelberg: Springer.
Trewavas, A. 2007. Plant neurobiology: All metaphors have value. *Trends in Plant Science* 12: 231–233.

Chapter 9
Computer Science Meets Evolutionary Biology: Pure Possible Processes and the Issue of Gradualism

Philippe Huneman

Abstract This paper investigates the relations between biological evolution and computer simulations of evolving entities through natural selection. It argues that what is proper to algorithmic evolution is that the selective dynamics of one modeled entity – for ex. genes, or species – is happening in the simulation with no immediate entangling with other levels of a hierarchy, unlike in biological evolution, where all the levels of the biological hierarchies are present together and their selective dynamics are entangled. This amounts computer simulation to propose "pure possible processes" of evolution, i.e. processes for which we know what kind and level of selection is at work. Algorithmic investigation therefore suggests processes as *candidate explanations for the patterns* of evolution we see out there. First, this fact allows one to solve issues which have been recently raised about the validation problem for simulation; second, in those conditions computer science is also likely to suggest new kinds of evolutionary processes whose outcomes would be *discontinuous* patterns of evolution. Drawing on recent work by Richard Watson, the last section of the paper will therefore consider how the longstanding issue of gradualism vs. discontinuities in evolutionary theory can be reassessed on the grounds of new insights provided by simulations like genetic algorithms. In conclusion I qualify the strong AL thesis according to which evolution by natural selection can be conceived of as an algorithm, and evolutionary biology as a branch of a general science of such algorithm.

Keywords Algorithmic evolution • Biological evolution • Computer simulation • Computer science • Evolutionary theory • Genetic algorithms

It seems that in many cases biology has to rely on simulation, because some of the most simple systems exhibit already features of non-linear causal relationships that preclude any analytical approach (Wagner 1999). On the other hand since a

---

P. Huneman (✉)
IHPST, CNRS / Université Paris I Sorbonne, Paris, France
e-mail: huneman@wanadoo.fr

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology, and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_9,

decade we witness an increase of studies in computer sciences that borrow terms and topics from evolutionary biology (Holland 1995; Mitchell and Forrest 1993; Forrest 1996; Ray 1992; Langton 1996, Hedrich 1999). A whole branch of computer science, mostly devoted to optimization algorithms but not exclusively, developed on the basis of insights from evolutionary biology: about genetic algorithms scientists speak in terms of genes, crossing over and mutation, and evolutionary computation (Koza 1992; Mitchell 1998) is a proliferating branch of computer sciences. Recent journals like *Bioinformatics* or *Biocomputation* exemplify this trend at an institutional level. The most biologically-oriented of those scientists, the Artificial Life practitioners, sometimes see their field as the investigation of life as it would be in any possible world – providing that we have a minimal set of criteria of life in terms of heredity and variation, which eventually resolves in a general susceptibility to Darwinian processes (Langton 1996; Bedau 1998; Lange 1996). In this radical sense, evolutionary biology becomes a part of computer science as science of general evolution, or, curiously, a kind of applied computer science – "applied" in the same sense than mechanics is somehow applied mathematics because a formal concept, namely differentiation, is therein physically interpreted. DNA here would be to pure evolutionary theory as a branch of computer science what speed is to mathematics. As implausible as it sounds, this claim has the virtue of being thought-provoking.

Whereas computer scientists often clearly refer to biology, biologists are less willing to consider computer science as such as relevant to evolutionary biology. An exception in the philosophy of biology is the famous work by Dennett (1995), which explicitly takes Darwinian selection as an algorithm, and evolutionary theory as the study of such Darwinian algorithm – a theory in fact mostly focusing on the genetic implementation of this algorithm, but not in principle wedded to this level. Another interesting occurrence of a serious consideration for algorithms is to be found in Maynard-Smith's paper on the specificity of genetic information (Maynard-Smith 2000); in this paper, he argues for the legitimacy of the informational talk in genetics, provided that information is taken in the sense of semantic information. Interestingly, his argument mainly relies on a parallel between biological evolution by natural selection and genetic algorithms.

In this paper I want to assess the validity of those hints towards an affinity between evolution by natural selection and algorithms, by investigating the relationships between computer science and evolutionary biology, in order to uncover what underlies the unity of their approaches. Shall we accept that algorithmic models are a generalised biology that deals with "possible" organisms as they would exist in any possible world? In this case, are the digital entities representations of those organisms (Langton 1996; Ray 1992; Forrest and Jones 1994), or are they themselves living entities (Adami 2002)? Or shall we only claim that computer science is an experimental extension of evolutionary biology, a tool for solving some of its issues – so that in this case we could not say that computer science and biology are two connected areas in a general science of Darwinian processes?

I will claim that one interest of the computational viewpoint for evolutionary biology, its ability to clarify null hypotheses in terms of outcomes of pure processes clearly defined in abstraction from other processes, through which as a consequence it may provide new models or paradigms of non-gradual evolution, thus bypassing the classical assimilation of evolution to gradual evolution through cumulative natural selection. Thus the first section will sketch some general distinctions relevant to the domain of computer modelling that will be useful to address the relations between evolutionary theory and computer sciences. For the moment I am rather vague on the term "computer science", given that it encompasses several kinds of algorithms studies: optimisation algorithms, evolutionary computation, and in general evolutionary algorithms, and then the more specific field of Artificial Life. The specification of those fields, as well as some of their articulations, and the extent to which they are concerned by this study of their relationships with evolutionary biology, will become clearer in the course of the paper. The second and third sections will formulate the idea that computational modelling devises a range of what I call pure possible processes that play the role of null hypotheses for evolutionary biology, and enunciate on this ground the validation problem regarding simulations. On this basis the fourth section will show that computer sciences provide evolutionary theory with a way to address the issue of gradualism, which is left unsolved in the Modern Synthesis.

9.1 Typology of Evolutionary Algorithm Studies

Addressing the wide domain of biologically-oriented computational models, two major distinctions have to be stated. The first one concerns the *context of simulation*, the second one is about the *uses of simulations*.

The first kind of context of simulation is what I call the "Formal selection context". Here, we assume that selection is a process unbound to its terrestrial realisation by nucleotides, amino acids and so on, which occurs as soon as we have replicating entities and variation among them. Simulation provides thus a context to observe the effectiveness of selection, but a selection with no living matter, hence a *formal selection*. This is in many ways relevant to life sciences, above all because such studies show the bearings and strength of selection in faster time given that evolutionary timescales encompass several decades, except for bacteria whose generation time is very short. Maley's study on biodiversity (2001), showing that emergence of new species is more conditioned by geographical barriers between populations than by adaptive potential, is a good example: several species are modelled, their members reproduce very quickly on screen, and then the researcher assesses the respective importance of the geographic specificities, and of some proper characteristics supposed to be related to adaptability (e.g. polymorphism, size, etc.): it appears after lots of generations that the first parameter is more explanatory – whereas in real time such studies would have required huge amounts of time. The simplest example in this category could be Maron's modelling of the oldest paradigmatic case of natural selection, Kettlewell's industrial melanism (Kettlewell 1956). He rebuilt moths on his computer, and the effects of selection by darkening (e.g. pollution) of environment occurred, in the short time of several

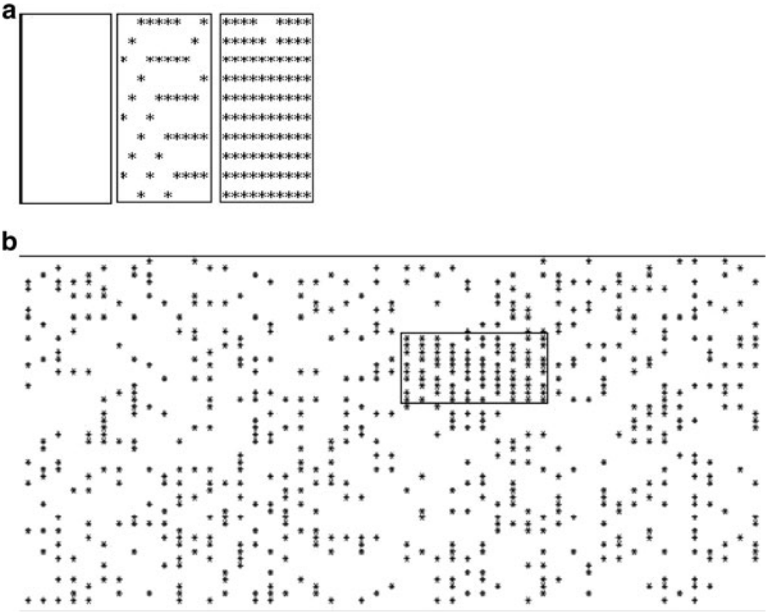


Fig. 9.1 (**a**) Three wing patterns. (**b**) A sample moth highlighted on a background with .25 density (Maron 2004)

hundreds of generation of those digital entities. Each moth is a 18 bits string. The first 8 bits encode a 2D cellular automata rule, while the 10 others encode the initial state. Running the cellular automata gives a wing pattern. Reproduction copies the entity near its location. Each predator and each prey have a given probability of feeding, death, birth. Those probabilities define a genetic algorithm on the cellular automata – namely, the moths. In a way similar to Mitchell / Crutchfield's results (Crutchfield and Mitchell 1992; Crutchfield et al. 1993), establishing that CAs can be evolved by a GA to the best rule for a given optimum, Maron's algorithm will evolve cellular automata towards fittest moths, namely the best at reproducing. It happens that their genotypes will match the environment in a way similar to Kettlewell's view of melanism (Fig. 9.1).

Given that natural selection *qua* selection increases the mean fitness of the population of elements it operates upon, the whole discipline of optimisation algorithms extensively relies on forms of selection, and hence provides cases of "formal selection context". However, the other context in simulation could be in opposite called the "no selection context". Kauffmann's work on "order for free" seems to be the best known example, since his models exemplify cases of the emergence of hereditary properties by physical process of self organisation, with no selection. Fontana and Buss (1994) conclude their enquiry on replicating macromolecules with those words: "our exploration of an abstract chemistry not only provides a route to generate such entities, but illustrates that different organization grades can arise in the absence of selection." All those investigations have been phrased by the term "self-organization". But some could be better understood as investigations of the prior conditions of a natural selection process, since, like models in Kauffmann's (1993), they investigate the articulation between selection and selforganisation.

Reynolds's *boids* are another interesting example (Reynolds 1987). They are digital entities in an agent-based model simulating the flight of herds of birds, through three simple rules implemented in each agent, concerning basically relationship to the distance and position of the other agents. Here the "building blocks" are simulating the birds themselves. There is no selection at stake, the point is to see whether with very simple rules that one can formulate by observing the flocking birds (for example: maintain a minimal distance with neighbours, tend to the centre of the group . . .) a flocking behaviour can emerge (which happens to be the case). However, those simulations are also accurate for school fish – whereas since the nervous systems of birds and fish are not similar, one cannot assume that the individual rules which really explain the behaviour will be the same across those two lineages. In another context, Mc Shea (1996, 2005) designed simulations of increasing complexity trends with no selection (see below) – that could match patterns of punctuated equilibria in phylogenetics. Also about evolutionary patterns (rather than processes), Chu and Adami (1999) designed simulations of phylogenies whose parameter is the mean number of same-order daughter families of a family; they have shown that for some plausible values of this parameter, and independently of what are the processes of evolution (be they primarily natural selection or not), power laws distributions for taxa of a same rank are expectable – e.g., clades with lots of species will be ten times less numerous than clades with ten times less species, clades with 100 times less species less numerous than . . . , etc. This justifies that sandpile distributions can be expected as outcome of evolutionary processes.

In those cases we must therefore make distinction between the *optimisation* algorithms, which will embody formal selection context, and some other cases of evolutionary algorithms which allow "no selection context" simulations. Orthogonal to this distinction, we have to draw another line, between *uses* of simulation. Humphreys (2004) initiated the methodical investigation of computational devices as ways of enlarging our epistemic capacities; he emphasised the fact that the usual categories of philosophy of science (experience vs. theory, laws vs. facts, observables vs. theoretical entities . . .) are not appropriate to account for the kind of knowledge delivered by those new devices. He made a useful distinction between simulations relying on previous theories about the field modelled, and simulations based on phenomenological rules that are grossly noticed as pertaining to some phenomena, and then implemented within the rules of the computational devices. Thus I will draw here a very close distinction, intended to capture features of the most biologically oriented computational devices.

Concerning natural selection there is a *strong* use, and a *weak* use of them. This distinction between strong and weak simulation is akin to Pattee's difference between realization and simulation (Pattee 1989), or Sober's distinction between weak and strong AL (Sober 1991), or even to the distinction that Bedau draws between models that "strive for maximal fidelity to the details" and "unrealistic models" in biology (Bedau 1999, p. 20).

The weak use consists in *testing some hypothesis* on selection. Miller and Todd (1995) modelled the sexual selection in this manner, demonstrating the numerous and often neglected effects of sexual selection on diversification, adaptation, exploration of adaptive possibilities : they assume that mate choice makes female internalize natural selection (they pick up males that are fit to environmental demands) so that what distinguishes sexual selection from natural selection as such is an advantage given to "exploration" behaviour by males – hence males and females are not playing the same role vis à vis natural selection, males are more exploratory and females devoted to "exploitation" of environment. The hypothesis predicts different roles in diversification and adaptive radiation, roles that the simulation displays. In the field of population genetics Maley's work, or Todd and Miller's work are good examples; simulations provide an extension of the testing capacities of models regarding hypotheses about dynamics. This could be called a *weak* use of simulations, since they are subordinated to the hypotheses and theories they are intended to test. On the other hand, the *strong* use assumes that "digital organisms" as "defined by the sequence of instructions that constitute their genome, are not simulated: they are physically present in the computer and live there" (Adami 2002), it is the world to which they adapt that is simulated.[1] So here, we do not simulate in order to test a hypothesis found when dealing with real populations, but we have a new kind of population ("new" in the sense that it has not to be understood in reference to a reality it is supposed to model), *about which* we can forge new hypotheses. Entities in those simulations are not necessarily simulating anything (e.g. cells, genes, etc.) in the real world. Those hypotheses made about them are capable of being more general claims about natural selection, because the peculiar, terrestrial, domain of its application is neglected. Strong simulations are not a mere way of testing claims, they are *in the same time* the domain of hypotheses and the field of testing. Ray's Tierra, Tim Taylor's Cosmos or Channon's Geb are good examples of this, and so are Conway's Game of Life and the extended rules of cellular automata in the Brain, Margolus or Life mode (see Margolus and Toffoli 1987). Of course one of the most famous is Holland's Echo, intended to be a simulation of ecosystems, with "organisms" that meet, fight or not, then lose or acquire resources in function of which they can reproduce at the next generation (Holland 1995; Forrest 1996; Forrest and Jones 1994).

In strong simulations, unlike weak simulations, the parameters included in the hypotheses are not given and can vary. For example, Langton's (1996) loop is a 2D cellular automaton in which emerges a loop which self replicates (Fig. 9.2), whereas the rules code only for the dots composing the loop. Thus we may not have in advance building block entities which by hypothesis are capable to replicate,

---

[1] On this claim, see Lange (1996). The essential point, as he claims, is not that a feature of life is imitated by the simulation, but that it gives an insight into the reason why this feature might be displayed. See p. 241 on flocking boids.

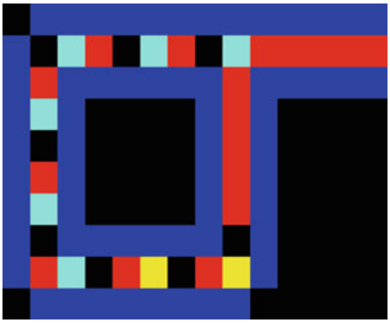


Fig. 9.2 The loop in Langton's loop CA

but the self-replicating feature is itself constituted in the course of the automata. Recently, Füchslin and Mc Caskill (2001) devised a simulation of molecules that evolve replicating templates and then a robust code, so that the simulation exhibits a robust origin of "genetic" coding. The oldest and classical example of emerging entity is the "glider gun" in Conway's Game of life. Thus, when we simulate we do not test assertions about self-replicating entities (such as in the Fontana-Buss example), but self-replication is an immanent property of the simulation (and does not by itself simulate anything). Furthermore, Sayama's (2003) SDSR loop,[2] an improvement on Langton's loop, exhibits differential reproduction – since some loops die while other reproduces – , so it defines a fitness function of its own, within the simulation itself.

John Holland's Echo has been used both in modelling natural phenomena such as host-parasite relationships, or mimicry. However the unrealistic assertions concerning reproduction and (in the case of mimicry) the absence of isolative reproduction between species, make those simulations poor in teachings for evolutionary biology (Crubelier et al. 1997). Therefore a richer use of Echo, – indeed, coherent with the notion of "strong" simulation – is to let it run in order to get ideas about potential evolutions of ecosystems, mainly about their stability, the patterns of associations, etc. To this extent Echo is a clear case of strong simulation

9.2 Natural Selection and the Pure Possible Processes

If we consider computer simulations as such and abstract them from what they simulate – *id est*, biological evolutionary processes – but consider them in themselves, as pure digital devices like genetic algorithms and cellular automata, I will argue that they are often experiments on the *modes* of evolution by natural selection. In this sense, although weak and strong simulations are different uses of them, they share some properties about the way they contribute to our knowledge of evolution.

---

[2] see also Sayama et al. 2003.

$a_1\ a_2 \ldots\ a_{1+P}\ a_{2+P} \ldots\ a_k \ldots$

$A_1^{n,j}\ A_2^{n,j} \ldots\ A_k^{n,j} \ldots$

$b_1^{n+1}\ b_2^{n+1} \ldots\ b_k^{n+1} \ldots$

Property P^n at Step n (Add (*on i*) Disj (*on j*) $A_i^{n,j}$)

Property P^{n+1} at Step n+1 (all the b_i^{n+1})

“If P^n had not been the case, P^{n+1} would not have been the case.”

Causation as counterfactual dependence between steps in CAs

Fig. 9.3 Causation in CAs (After Huneman 2008b)

First, in some sense, computer simulations as such involve causal relationships and processes. Let’s establish briefly this point, in order to understand what those simulations can bring to the study of causal processes in evolutionary biology. Suppose a cellular automaton with cells a_i at step n. To show it, let’s call a_1^{n+1} the state of the first cell on line $n+1$. Let’s call A_1^n a set of states $a_{1+m,\,n}$, m varying from 1 to p (p is defined by the rules) such that their outcome at level $n+1$ is $a_1^{n+1.}$ Now, there is j other sets of states like A_1^n such that the outcome is still $a_1^{n+1.}$ So we can index those sets of states: $A_1^{n,\,k}$. The property P is then defined : a CA is said to have property $P^{n,1}$ at step n iff it exists k, $0 < k < j+1$, such that it is in a set of states $A_1^{n,\,k}$.

Now for any i, and the state of a_i at step $n+1$ we can define a set of states $A_i^{n,\,j}$ that all result in a_i^{n+1} and thus define the property $P^{n,i}$. And let’s define Q the property of being in the state $\{a_1 \ldots\ a_i \ldots\ am\}$, the property instantiated by the CA at step $n+1$. Finally, we can say that the CA at step n has the property P iff it has all the $\{P^{n,i}\}$. Then, if the CA had not been in a state P at step n it would not have been in state Q at time $n+1$: counterfactual dependence, hence causation. So the only causal explanation of “being in state Q at $n+1$” is “being in state P at n”. In computer simulations like cellular automata there are causal relations between sets of states at various steps. I argued earlier in Huneman (2008a, b) that a subclass of these automata display a specific type of structure which allows one to talk of emergent processes, and that (Huneman 2008b) within those emergent processes some possess also a causal specificity, which is the emergence of causal relations of higher level than the elementary causal relations (Fig. 9.3).

Therefore, in a computer simulation like cellular automata there are not only formal but also counterfactual dependencies between some states of cells, i.e. causal relationships. This establishes that computer simulations will be likely to provide tests for causal hypotheses in biology; as I will elaborate it from now on.

Since in formal selection contexts models deal with natural selection, states are compared in terms of their differential reproduction and persistence. The causal processes, here, are of selective nature, and those models embody rather than simulate natural selection. Finally, strong simulations appear to be a general investigation of some causal processes, of which natural selection is a prominent part. This investigation can be said "experimental", when no equations can be designed – experiments here in the sense of computer experiments (see Winsberg 2003, see also Humphreys 2004 on non-theory-based computational devices) – but may be theoretic when analytic treatments of the dependencies within the algorithm can be provided. Weak simulations are testing hypotheses because they equally produce causal processes – selective processes in the case of formal selection models – to check whether these are indeed likely to produce the outcome which is our explanandum.

Those selective processes – present in the two uses of simulation – however are of various kinds, that computer scientists intended to characterize. They drew a distinction important for our purposes. Channon and Damper (2000) termed "artificial selection driven" the GA systems where evaluation of the fitness function is fixed in advance, and "natural selection driven", the systems where one cannot make a finite specification of what it is to be fit. Note that this second case is what occurs in natural ecosystems, where the same fitness value could be realized by (for example) either a high race speed or a perspicuous sight – so the labels "natural vs. artificial selection" here are not simply metaphorical.

Some systems, in simulations, are designed to undertake Channon's "natural selection". Ray's Tierra (1994), Hillis's coevolution (see Hillis 1998) are well-known examples. Here, the fitness function is not determined in advance; and we see that a population can evolve some new features, since they can be conceived as new classes of entities, and then behave in a specific manner in the simulation. Hillis improved an optimizing GA for sorting networks by creating another population, anti-networks, namely "sorting problems"... Those problems are intended to minimize the fitness of the sorting networks, because they randomly display difficulties for which those algorithms have not been designed by selection; hence the fitness function itself evolves according to the arising of sorting problems. A dynamical coevolution analogous or identical to host-parasites coevolution occurs. Similarly, in his population of Tierra programs, Ray has found arising "parasites", "hosts", "hyper parasites" and so on... "Ecological" adaptations were therefore going on. Later, Channon designed a system called Geb, able to perform emergence of innovations. Nevertheless, he recognizes that a limit of those simulations is that "specifying the available actions constrains organisms around these actions and so limits evolution" (Channon and Damper 1998). Even if the fitness function is likely to be left open, the class of actions of which building blocks are capable is finite.[3]

Holland (1995) has shown that in the Echo model, parasitism and some other features of the ecological settings are likely to appear, for example instantiations

---

[3] This might in the end proves to be the main difference between simulations and the biosphere.

of the competitive exclusion principle. However an important difficulty, for one who would consider those Echo experiments as parts of ecology, is that they don't provide univocal definitions and criteria to single out the genuine *species* in the Echo simulations (Forrest and Jones 1994): depending upon the choice we make of considering a genotype (i.e. one agent), or a set of agents with various genotypes (collected under some principle), as constituting a *species*, the interpretations of an Echo round will strongly differ.

In this perspective, I make two remarks that stem from the above considerations. (a) it is clear that when we get to the open-fitness computational systems, the relationship between computer science and evolutionary biology becomes *much more than the analogy that was supposed to hold in the case of weak simulation*. What occurs among the entities, be they cells in a CA, programs in a GA or "agents" in an agent-based models, is indeed a genuine causal process, the kinds of which are also the *instances of natural selection in biology*. Strong simulations are indeed likely to *embody the same processes* as biological evolutionary phenomena do. Of course, in one case the underlying realities are physicochemistry, whereas it is digital entities in the other cases, but the *processes* undergone by, on one side chemical entities, on the other side digital entities, can be equated. The vocabulary of crossing-over, mutations, genes, etc. in Genetic Algorithms, for example, is not metaphorical, those processes are really occurring here, *but* are undergone by another kind of entities (non carbon-based genetics, if you want). From this viewpoint, evolutionary theory is not something imported in computer sciences, but those investigations in algorithmic modelling that we are addressing here (formal selection strong simulation, above all, because they abstract away from contingencies of this-wordly evolutionary entities) are part of a generalised evolutionary theory. Computer sciences and biology then appear as two branches of the same tree, like condensed-matter physics and physics of liquids.

However, (b) the *entities* engaged in those computer processes – unlike the processes themselves – are not easily mapped onto biological entities. This is obvious in the case of strong simulations: species are difficult to be isolated in Echo, lineages are not easily seen in Tierra, etc. Besides, unrealistic assertions (Bedau 1999) are often at the basis of computer simulations (absence of epistasis, of linkage disequilibrium, etc.), which compels one to be greatly cautious while interpreting their results in biological terms. More generally, even if some computational systems are designed to correspond to *one* set of entities – let's say the species, in order to study phylogenetic patterns (Chu and Adami 1999; Mc Shea 2005) – the *other* levels of biological reality (organisms, chromosomes, etc.) are not in the same time given and identifiable in the systems.

This fact exemplifies something general about computer simulations, that is crucial to their epistemological status, and that I explain now. Only in real biology one can trace in a same biological setting (within a population of one or various species) connections between species, organisms, chromosomes lineages, or generally microevolution and macroevolution; only in real biology are the whole evolutionary hierarchies, *together* ecological (genes, cells, organisms, populations, species) and genealogical (genes, organisms, species, genera etc.) (Eldredge 1985), really given *in the same time*, in the sense that when you have cells you have both genes (the level below) and organisms and species. Yet if one computational system is designed to correspond to some level of biological reality, the other levels are not *ipso facto* given (whereas if you have organisms you have genes and species).

A major consequence of this is: while phenomena in biology might pertain to *various* processes involving *several* of those entities, and sometimes entities not taken into account in the biological models, within the computer simulations one can't meet processes that would involve types of entities that are not capable of being defined in this simulation on the basis of the explicitly designed entities, i.e. on a simple, univocal and determined relation to the building blocks (for example, cells in a CA). Causal processes in computer simulations can't involve entities that don't algorithmically exist, while the entities undefined in biological models still biologically exist and biologically interact with the ones we have modelled. A paradigmatic case is multilevel selection: scholars increasingly acknowledge the fact that natural selection plays at several levels (genes, organisms, possibly species or clades), and in possibly opposite directions; this multilevel process is then responsible of many aspects of biology (cooperation, mutualism, plausibly the emergence of organisms multicellular and societies, etc. (Michod 2005; Damuth and Heisler 1988; Okasha 2006, etc.)) However, suppose that I devise a simulation with "genes" as entities: nothing guarantees that there will be the possibility to define species or cells in this simulation, so if selection acts here, it is clear that there will be only selection at the level of genes. This fact is crucial to understand the epistemological value of computer simulations, as I will explain it in the next section. And obviously, this might account for the limits met by the computer systems when one wants to directly simulate ecological phenomena in Echo (Crubelier et al.), or to model long-term trends in evolution (Bedau et al. 1998; Bedau 1999 vs. Mc Shea 2005).[4] Notice that I do not contend that in a simulation there is only one class of entity, because I won't preclude the possibility that some higher level entity can emerge (in fact lots of those simulations are intended to cast a light of this phenomenon of emergence); but my point is that nothing guarantees that all the levels are definable in one given simulation, unlike biological reality where all the levels of the hierarchies coexist, so that the causal process proper to all those levels do occur in the same time.

Whereas (a) supports arguments of those who claim that evolution is an algorithmic process, or that evolutionary biology is a branch of a general evolutionary computational theory, (b) implies structural differences between processes in

---

[4] Mc Shea (2005) indicated that if we define complexity in non-functional terms, internal variance in life (due to the mere principles of heredity and variation) provides a trend toward increasing complexity (see below). This trend is matched in some parts of the known phylogenetic tree; however this process seems like diffusion in physics. Bedau (2009) argues against this idea by saying that what is increasing in this diffusion process has not necessarily the meaning of biological complexity that Mc Shea intends to study. Something is increasing, but the increase due to this diffusion has no biological meaning that could be identified within the simulation.

evolutionary biology and processes in simulations, a fortiori in strong simulations, differences that undermine the strong thesis of identity held for example by lots of AL practitioners.

The next section draws the consequences of those facts as to the validation of computer simulations.

9.3 Facing the Validation Problem: The Variety of Selective Processes in Computer Sciences and Their Impinging on Evolutionary Biology

In this sense, the value of algorithmic research for evolutionary biology is that they provide pure versions of some classes of processes, and an idea about what they are capable of producing. "Pure" here means that they are unmixed with other processes that could go unnoticed in genuine biological cases, because the entities involved in those parasitic processes, which in biology are accompanying the focal entities, would make radically no sense in the algorithmic universes considered. The epistemological fact here underlying the notion of pure processes is the "transparency" of computer algorithms, taken in two senses: the mechanisms involved being defined by us, nothing else than them is responsible for the outcome (Bedau 1999); if something new emerges, we know what it emerges from. Now, in the case of formal selection simulation, where the processes pertain to natural selection, those models display pure selective processes concerning the entities modelled (e.g., pure species selection).

The question of whether those classes of processes match actual biological processes has to be raised in a second step. It arises, precisely, when one compares the *patterns* of evolution found in nature, in the fossil records, to the outcomes of such classes of *processes*, and notices that one of them maps onto a computational outcome. This is not as such an evidence for the occurring of such a process, but it is a suggestion that this kind of process is likely to be the genuine explanation. However, this assertion further requires that the simplifying assumptions made in order to design the simulation – the definition of the agents, the ranges of their actions, the way they combine and replicate – are not too unrealistic. It requires also thinking about which processes might have occurred simultaneously in nature, since in simulations we have only the single process that we intended to simulate. So if, for example, a phylogenetic pattern exactly matches the result of the simulation of species selection acting on a set of species realistically defined, this does not prove that such pattern results from species selection since other processes were acting in nature.

Hence simulation experiments inform about what classes of processes are *likely* to be found out there – exactly in the same way as logics informs us about what *can* be (and be said), given which phylogenic patterns we have found in nature. Hence what might be the proper input of computer sciences in evolutionary theory is the connections made between *patterns* of evolution and *classes of processes* (mainly selective processes). They cannot *prove* that a process with such and such parameter accounts for this or that feature of evolution, but they allow us to figure out *candidate explanations*. Their epistemic value consists only in pointing out what kind of evolutionary process can be at stake, *even if the selective pressures are not identified* – which should be the proper study of ecology. The idea is: *if pattern X is met, then process x is likely to have produced it*.

In a paper on complexity in the genome, Lenski et al. (2003) show that evolution is likely to have favoured complexity through episodes of decreasing mean fitness. Their point is that, if there is complexity increase (in the sense of complexity chosen in the paper), then deleterious mutations might have been selected; thus a decrease in fitness might have been involved in the stabilisation of more functionally complex genomes: the simulation displays that rearranging the genome after a functional innovation is shown to involve a temporary less-functioning episode, hence a decrease in fitness. Thus, the general modes of getting increasing complexity are illuminated by the simulation, while no teaching is provided concerning what actually happened. Similarly, in Chu and Adami's (1999) investigation of the patterns of abundance of taxa (cited above), the power-law distributions that they got as outcome provide ideas about what would be the causes of a phylogenetic pattern: if the distribution of taxa resembles a certain power-law scheme X, it is likely that the parameter m (mean number of same order daughter families of a family) has been in nature close to the value of m involved in X when X results from a simulation with such value of m.

In those examples we face a classic epistemological puzzle about computer simulations, that one could call the *validity paradox*: what tells us a simulation whose outcomes match reality? Epstein (1999) famously gave the example of the historical expansion and decrease of the settlements of Anasazi Indians. Simulations have been made, in terms of agent-based models whose agents are Indians and whose actions are resources foraging, resources sharing, neighbours fighting etc. When you run the simulation, with some initial distribution, you get in the end a geographic pattern quite identical to the real pattern described by archaeologists. Epstein claims that this confirms the simulation's rules as hypotheses about the causes of this situation: the Anasazi acted in such and such way, implementing the rules enacted by the simulation. However, as Grüne-Yanoff (2009) emphasises, this does not *prove* that some alternative explanation, ascribing very different major actions and motives to Indians, would not yield the same outcome, so it does not rule out all alternative explanations, hence it does not prove the genuine motives. Moreover, Küppers and Lenhard (2005) argued that the epistemological validity of simulations consists rather in the efficient matching of the outcome with real pattern, and hence the possibility of predicting patterns, than in the correspondence between the rules of the agents in the models and the real-life agents' rules. Their argument referred to Arakawa's meteorological simulations, which reached a very successful matching with genuine weather patterns, much better than previous simulations, but at the price of making even more unrealistic the agents' rules and the boundary assumptions.

To this extent, computer simulations in evolutionary biology cannot be said to prove anything concerning *real evolution and the genuine causes of evolution.* I will briefly here mention two examples.

Think about Reynolds's flocking Boids. For sure, the final outcomes match the real flight pattern of a flock of birds, or even a school of fish; yet this does not prove that the rules implemented in the nervous systems of the birds – and of all birds, irrespective of their species – are the three rules to which the agents obey in the model. So what does it prove? It proves that the pattern of group flight does not *need* any central control to be realized – so it somehow excludes any social organisation process as an explanation of the group flight. But note that it *truly* circumscribes the set of plausible hypotheses *only if* one subscribes to a value of *parsimony* that is implicit here, and that entails that central-organisation explanations are less parsimonious than agent based explanations.

My second example – which is more relevant to the science of evolution as such – is Mc Shea's work on complexity (1994, 1996, 2005). In a series of papers, this author designed models of what a complexity trend would be in evolution. At stake is the question of the nature of any trend in evolution: passive or driven, and caused by what? Mc Shea distinguishes those two questions in terms of mechanisms (passive vs. driven) and causes (natural selection? diffusion? etc.) of the trend (Fig. 9.4) . He builds computer simulations where basic entities are the species, which have a fixed probability to speciate, increase in complexity or go extinct. Simulations model either cases with boundaries (e.g. a lowest level of complexity) or no boundary. Computer simulations here allow us to find out that the mere fact of variation (the variance of the value of a character increases generation after generation even if a mean value is stable) produces a trend toward increasing complexity, even in the absence of natural selection (Mc Shea 2005). In the simulations, we also state what would be either a driven trend or a passive trend in evolution, so that we can confront the phylogenetic data in order to test which kind of trend (if any) is to be met (Mc Shea 1996) – the answer being a driven trend. The first result decouples issues of increase in complexity from the process of natural selection, allowing thus biologists to overcome the schema of complexity increasing with size through natural selection that Bonner (1998) explored; the second result shows what evolution would be like if some kinds of trends (namely, driven trends with no selection as a cause) were present within it.

In those cases, algorithmic investigations set the ground for explanations of trends in evolution – and above all, complexity trends – so that one knows what are the minimum conditions that make increasing complexity expectable, and also knows how to recognize various kinds of trends (passive vs. driven), which eventually allows a choice between the possible causes of trends. Interestingly, the simulation shows something about the phylogenetic pattern of complexity distributed across species, namely that some trend may have been at work here. It suggests that this trend is as such explanatory, with no reason to enter into more fine grained analysis. A parallel situation occurs in ecology with the so called ecological neutral theory initiated by Hubbell (2001). In those models a fixed probability of death and birth is ascribed to individuals of each species; yet

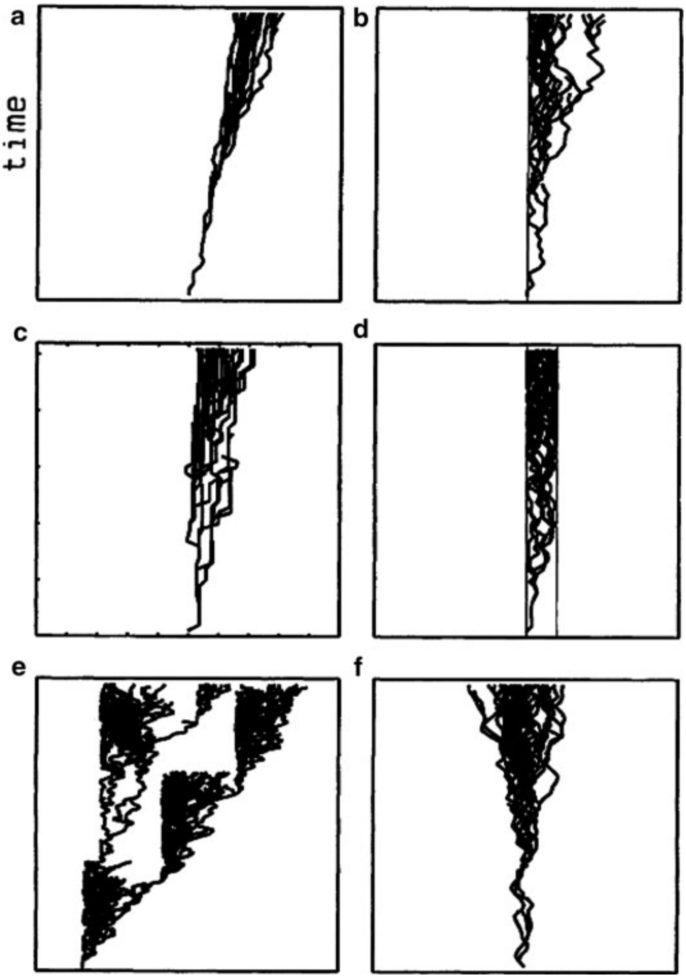


Fig. 9.4 Mc Shea's trends in complexity in evolution

the occurring distribution of species in modelled communities, as well as their succession, match pattern of distribution within real communities. This suggests that whatever the selective processes are at the level of individual organisms, species distributions within communities, and their succession rules, behave like a random walk. In Hubbell's model actually the assumptions about birth and death rates are wholly unrealistic about individuals since they amount to denying natural selection (i.e. differential reproduction). The simulation here does not provide the explanation but indicates that some processes (here, species level purely stochastic interactions disregarding what happens at the organisms' level) are likely to produce by themselves the actual pattern of species distribution.

In this sense, as we can see from those three examples, the simulations, even when their outcome match the empirically detected trends, do not explain them, but *forge candidate explanations and compare their likelihoods.* This latter term is taken in the technical sense of the probability of data conditional upon the hypothesis: here the probability of a phylogenetic or ecological pattern conditional upon some hypothetical causal process. Moreover, in many cases they provide some kinds of null hypotheses: for example in Mc Shea (2005) for phylogenies, or in Hubbell (2001) for ecological communities, they sketch "what would occur with no selection", In this respect, the epistemological value of algorithmic devices for evolutionary biology consists in providing null hypotheses which allow to classify and single out the various outcomes of the hypothesised processes likely to be at stake in the evolutionary reality.

This value is nowhere more obvious perhaps than when it comes to the issue of gradualism in evolution, onto which recent advances in bioinformatics or evolutionary computation have cast a new light.

9.4 The Issue of Gradualism and Computer Simulations

9.4.1 Neodarwinism and Innovations: A Longstanding Concern

There is an implicit equation in neo-Darwinism between natural selection and cumulative selection. Darwin insisted on the fact that variation are small, so that natural selection is slowly shaping the adaptations, and later Fisher and the Modern Synthesis refuted objections according to which small variations could not give birth to big novelties (Mayr 1976; Gayon 1998). However, this made novelty – as the emergence of a real qualitative innovation, e.g. fish gills, mitochondria, sexual reproduction, language – into a riddle for Darwinism. Two issues have to be distinguished here: issues of *complexity*, and issue of *innovation*, those two properties, while often conjoined, being conceptually distinct (Mc Shea's work considered in the previous section investigated only the latter). Granted, complexity is successfully described by classical neo-Darwinism, from the times of Darwin who devoted a chapter to a hypothetic history of the eye, to the recent adaptationist understanding of complex organs (Williams 1992; Dawkins 1982). Note that this issue of complexity has been enriched by perspectives from the evolution of digital organisms, through both the measures of complexity (Adami et al. 2000) and the proofs of an increase in complexity (Wilke and Adami 2003; Mc Shea 1996).

Computer simulations enabled one to put flesh on Darwin's speculation since we have now a simulation of the emergence of eyes which corresponds to Darwin's idea of intermediary species having each in its turn an increasingly efficient and discriminating light detector, and provides a plausible course of this evolution

(Nilsson and Pelger 1994[5]). In a systematic attempt to bring together insights of evolutionary computation and evolutionary biology, Wagner and Altenberg (1996) established that a condition for evolving complex adaptations is that systems are evolvable, which is mostly obtained through specific relationships between genotype and phenotype ("genotypes-phenotypes maps"): relations that prevent big local rearrangements of phenotypes to induce disrupt phenotypes, or in other terms "modularity". But innovation is not identical to complexity, it is properly speaking the arising and maintenance of a variation that was not behaviourally or morphologically included (*id est* present in a more or less accentuated form) in the previous traits. *Innovation* as such is a puzzle for Darwinism (Cracraft 2000), facing the issue of the non-adaptive value of small, non functional, incipient forms of novel traits: what would be the benefit of having, let's say, the tenth of a wing? – but this inchoative wing must yet have been selectively retained, if the whole wing was to appear... The usual solution, the one privileged by Mayr (1976) in his attempt to wholly understand innovations from the viewpoint of the Modern Synthesis, is that little steps in phylogeny happened, each displaying a partial state of the trait considered, and each having an adaptive value, exactly like the case of the eyes according to Darwin and later (Nilsson and Pelger 1994), or in the case of insect wings which in their inchoative states have been adaptations for thermoregulation, before being recruited for their flying capacities (Kingsolver and Koehl 1985).[6] Yet this option is not always possible, especially when it's hard to understand the adaptive value of a part of the trait (think of mitochondria).

Evolutionary theory of development, or Evo-Devo, has been massively interested in confronting this issue (Müller and Wagner 2003; Newman and Müller 2005; Müller 2007; Carroll 2005), sometimes giving up population genetics in favour of self organisation theories (Kauffmann 1993; Godwin 1996), that make the wholly structured innovation emerge at once on the basis of singular physicochemical properties. Interest in developmental features as major contributions to evolutionary *processes* has been fostered by the famous alternative approaches of the *patterns* of evolution, like the famous theory of punctuated equilibria, put forth by Eldredge and Gould in the 1970s (Gould and Eldredge 1977). Punctuated equilibria theorists claim that the lack of transition forms in the fossils record is not due to the geological constraints, but attests that evolution is a two steps process, alternating very long phases of stasis – where the main morphologies, body plans etc., remain essentially the same through the development of detailed adaptations -, and brutal (evaluated in geological times, of course) phases of evolutionary change, where new phyla, body plans, etc., emerge (Fig. 9.5). For example, the Turkana Basin displays fossils from mollusc lineages that are very similar, and groups of fossils indicating quickly evolving forms (Williamson 1981). While this theory is not tied to any particular process of evolution, evolutionists (following Gould 1977) mainly concentrated on the possibilities of recomposing the whole genotype in

---

[5] See simulation at http://www.blackwellpublishing.com/ridley/a-z/Evolution_of_the_eye_b.asp.

[6] Such changes have been famously called "exaptations" by Gould and Vrba (1982).

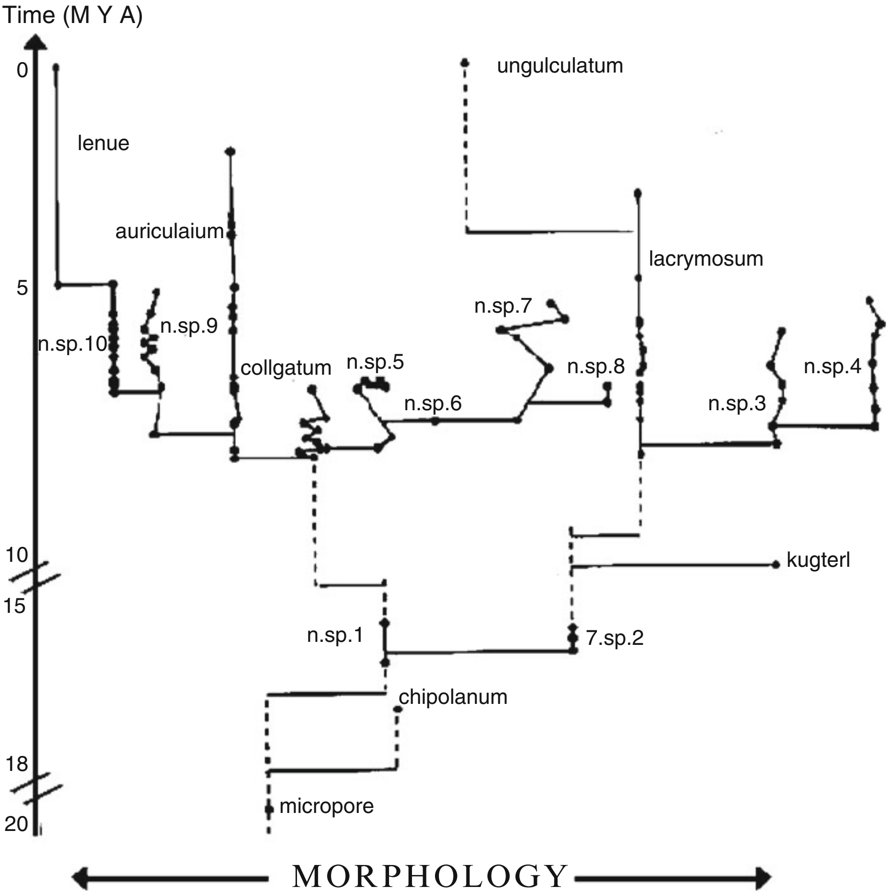


Fig. 9.5 Punctuated equilibria in the fossil records

a holistic manner through changes in the developmental timing (Müller 2000; Müller and Newman 2005; Carroll 2005; Raff 1996 among many others). Speciation seems to be accompanied by ruptures in developmental homeostasis, which in turn correlates morphological stasis with developmental homeostasis (Williamson 1981, p. 215). The emphasis on modes of development instantiates the connection between genotypes and phenotypes, i.e. the genotype-phenotype map, which accounts for differences in evolvability (according to Wagner and Altenberg, 1996). Yet the sudden evolutionary bursts proper to punctuated equilibria, while challenging gradualism, do not compel one to conceive of only one class of processes as responsible for them. At this point computer science is offering new perspectives on this issue too.

9.4.2 Compositional and Gradual Evolutions

In *Compositional evolution* (2005) Richard Watson investigated some not-yet-formalized modes of evolution by natural selection. The term names "evolutionary processes involving the combination of systems and subsystems of semi-independently preadapted genetic material" (p. 3). The interest of this approach is that the consideration of building blocks obeying some new rules that are inspired by the biological phenomena of sex and of symbiosis proves that in those processes non gradual emergence of novelties is possible.

The basic idea is that in terms of algorithms, interdependencies between parts of a system correspond to kinds of possible evolutions. (1) A system with weak interdependencies between parts can undergo linear evolution: increases in complexity are linear functions of the values of the variables describing the system (for example, the if the complexity is the number of cell types, one could call linear a complexity increase where this number follows the size of the system, its volume, its number of nucleotides, etc.). Algorithms looking for optimal solutions in this way are called "hill-climbers"; they are paradigmatically gradual. They easily evolve systems more complex in a quantitative way, but they can't reach systems that would display innovations.[7] (2) If you have arbitrary strong links between the parts, then evolving a new complex system will take exponential time (time increases as an exponential function of the number of variables). Here, the best algorithm to find optimal solutions is the random search. (3) But if you have modular interdependencies (encapsulated parts, etc.) between parts, then evolving new complex systems is a *polynomial* function of the variables. (Watson 2005, pp. 68–70) Algorithms of the class "divide-and conquer" are dividing in subparts the optimisation issue, and divide in turn each subpart in other subparts : the initial exponential complexity of the optimisation problem approached through random search is thereby divided each time that the general system is divided – so that in the end the problem has polynomial complexity. Those algorithms illustrate how to evolve systems that are not linear improvements of extant systems; but as polynomial functions of the variables, they are feasible in finite time, unlike random search processes.[8] So with a system of this third sort, patterns or features which would not be reachable by evolution in realistic time become reachable. This framework indeed concerns directly evolutionary innovations, because one can assume that most of the time gradual evolution (accumulation of selection on small variations on genes) relies on linear improvements of extant systems, hence seems unable to explain those innovations.

---

[7] Notice that fitness function, here, seems fixed in advance – we have mostly algorithms that embody the "artificial selection" in the sense of Channon and Damper cited above.

[8] This is somehow the main line of Bonner's approach of complexity, since he correlates complexity, number of cell types and size through natural selection, which is supposed to gradually improve size.

"Compositional evolution" concerns pure processes that embody those classes of algorithms with polynomial rates of complexification, and have genuine biological correspondents. After Lynn Margulis (Margulis and Sagan 1986), biologists became indeed convinced that our mitochondria initially were symbiotic bacteria which got encapsulated and eventually fixed in the nucleus. Acquiring an innovation through symbiosis obviously skips the riddle of the selective value of inchoative stages in innovations, since it makes no sense to ask what was the value of having, for example, half a mitochondria: the symbiotic part by definition evolved as a whole . . . Symbiosis easily exemplifies the "mechanisms that encapsulate a group of simple entities into a complex entity" (Watson 2005, p. 3), and thus proceeds exactly in the way algorithmically proper to polynomial-time complexity-increasing algorithms like "divide and conquer". Watson defined for genetic algorithms precise ways of evolving that correspond to symbiosis in quite the same way as the classic operators (mutation, crossing over, and replication) were defined by Holland (1995) in correspondence with the genetic processes of the same name.

Granted, crossover is usual in GAs, and embodies one feature of biological sex; it allows then for selection acting on blocks rather than alleles alone, which is a way to get modularity in GAs – modularity being often a key to evolvability (Wagner anf Altenberg 1996). Yet Watson refined the usual crossover clause in GA, integrating various algorithmic devices (for ex. "messy GA", according to Goldberg et al. 1991) in order to allow a selection on blocks which takes into account correlations between distant blocks, hence creation of new blocks (Watson 2005, p. 77). To this extent, the evolutionary potential of sex is taken into account in a more fine-grained and encompassing way here than in the usual GA research tradition – hence, the possibility of understanding necessary non-gradual evolution and the arising of innovations as outcomes of this evolution. Now, all those "encapsulation" genetic algorithms, no matter the rules that provide the encapsulation, are of the kind of polynomial-time evolutionary algorithms. This proves that biological processes formally structured like those encapsulated processes – such as symbiosis, endosymbiosis, may be lateral gene transfer – have been likely to provide evolvability towards the innovations not reachable through gradual evolution.

Watson (2005) has thus shown that cumulative selection as such cannot do better than a kind of linear increase in the value of some parameters, which makes difficult to understand how some innovations could appear through such evolution in finite time. Hence evolving innovative systems in a gradual way is not plausible (at least, as a general solution, though scenarios like exaptations and theories focusing on developmental variations (e.g. Kirschner an Gerhardt 2005) are explanatory here). However, adding the symbiotic and the sexual modes of evolution by selection makes available polynomial increases in the values of variables, which can account for the genuine arising of innovations that occurred in nature. This increase rests on a compositional process that would for instance embody the discontinuity in evolution proper to the acquiring of mitochondria, namely a sort of evolutionary jump corresponding to the addition of the already evolved symbiotic digital organism.

So the longstanding issue of gradualism is here addressed through a conceptual move back and forth between evolutionary biology and computer science[9]: biology provides a natural model for a kind of evolution which seems to display discontinuity (e.g. symbiosis), and hence speeds up gradual evolution; and in return computer science devises a pure kind of process – called "encapsulation" – that picks out the feature responsible, in the symbiotic evolution, for the speeding up of the process (namely, drops from exponential to polynomial time). The bulk of Watson's demonstration is the identity between algorithmic classes (hill-climbing, divide-and-conquer, random search) and evolutionary processes (gradual evolution, compositional evolution) – an identity which plainly makes sense in the light of the first sections of this paper. Finally, evolutionary theory, when it considers cases of evolution which are controversial on a gradualist viewpoint – because cumulative selection seems too slow to account for them – is provided with a process-model likely to explain such episode. So the solution of the gradualism issue is neither a quest of non-darwinian explanation ("order for free", etc.), nor a reassertion of the power of cumulative selection that needs to be more deeply investigated (Mayr 1976), but the formal designing of new modes of selective processes, of the reason of their specificities, and of the differences between their evolutionary potentials. In this sense, discontinuities in evolution appear as the explananda of a variety of selective processes whose proper features and typical evolutionary patterns are demonstrated by computer science.[10] For this reason Watson's work seems paradigmatic of the interplay between evolutionary theory and computer science, provided that – as it is claimed here – computer simulations are a way to provide candidate process explanations for actual patterns, and null hypotheses.

Further reflexion is needed to make sense of the *discontinuity* which is proper to novelty (Fontana and Schüster 1998). The work surveyed here provides anyway a basis of an attempt to relate and systematise the various processes of which computational models provide the pure outcomes. Discontinuity can be conceived of in terms of emergence, as opposed to adaptation continuously produced by cumulative selection (Huneman 2008b). In effect, emergence in computational systems can find a criterion, namely the unavailability of an outcome except by simulation (Bedau 1997, 2003; Huneman 2008a; Humphreys and Huneman 2008). Discontinuity and emergence characterize precisely the evolution of new fitness units, which is the target of the research program of "evolutionary transitions", initiated by Maynard and Szathmary (1995) and Michod (1999). It is patent that "compositional evolution" as construed by Watson is likely to address cases of evolutionary transitions (Watson 2005, p. 54), because its *encapsulation* model might generate in a more feasible and available way than gradual evolution (with its linear progression) those "entities that were capable of independent replication before transition (and) can replicate only as a part of a larger whole after it"

---

[9] For a table of correspondence between the two fields, see e.g. the table in Watson 2005, p. 42.

[10] As Watson himself says, computer science provide "means . . . to identify differences in adaptive processes in a fundamental sense" (2005, p. 2), i.e. in terms of rhythm and feasibility.

(Maynard and Szathmary 1995) and that are the target of the investigation in the "evolutionary transitions" program. So once again a radical case of evolutionary innovation can find in computer science the construal of pure possible processes likely to explain it.

The fact of discontinuity will provide a basis for a classification of the evolutionary patterns resulting from the various selective processes formally investigated by computer sciences – and therefore, a basis for answering the question of the processes responsible for increases in complexity in life, in the various senses of complexity Huneman (2010).

9.5 Conclusion

The Strong AL claim is: biology is a part of a very general science of evolution, instantiated by some branches of the computer sciences. This paper tried to assess such a claim, mainly by focusing on the relationship between algorithmic modelling and genuine cases of natural selection, and by understanding what epistemological value computer simulations have for evolutionary biology.

The main result is that computational models are not a very general domain of which biology would exemplify some cases. On the contrary they mostly provide pure possible processes that might causally contribute to origin of traits or evolutionary patterns. The class of possible processes being larger than the real processes, it is obvious that not all processes simulated – not all cases of artificial life, if you prefer – are likely to be met in biology. Most of the writers are accounting for this in terms of the chemical nature of living entities, which will provide boundary conditions for those processes. Yet I argue that the right account is given by the fact that biological processes are never pure, since (Sect. 2) in biology all entities are given together in their hierarchical scales, while algorithmic devices only permit to single out one or few entities within them. In this sense they are only generating the pure processes involving solely those entities.

To this extent strong simulations do rather provide general null hypotheses for evolutionary biology – in terms of taxonomy of pure processes, mostly (but not always) selective ones. This also settles the validation paradox for biologically oriented algorithmic investigations. One major interest in this virtual exploration of selective processes is that realistic challenges to gradualism can be formulated, in terms of pure possible processes, computationally modelled, likely to feasibly generate innovations. This allows more plausible and pluralist schemata of macroevolution, which contribute to theoretical integration in biology: integrating micro- and macro- and mega-evolution (sensu Simpson 1949); and integrating distinct approaches of innovations, from classical evolutionary theory (Mayr 1976) and from recent trends in Evo-Devo (Müller 2007; Gilbert, Opitz, Raff 1996).

Acknowledgments I warmly thank Laura Nuno de la Rosa Garcia, Sara Franceschelli and above all Anouk Barberousse for their careful reading and criticisms of this paper.

References

Adami, C. 2002. What is complexity. *Bio Essays* 24: 1085–1094.

Adami, C., C. Ofria, and T. Collier. 2000. Evolution of biological complexity. *PNAS* 97(9): 4463–4468.

Bedau, M. 1997. Weak emergence. In *Philosophical perspectives: Mind, causation, and world*, vol. 11, ed. J. Tomberlin, 375–399. Oxford: Blackwell.

Bedau, M. 1998. Four puzzles about life. *Artificial Life* 4: 125–140.

Bedau, M. 1999. Can unrealistic computer models illuminate theoretical biology. In *Proceedings of the 1999 genetic and evolutionary computation conference workshop program*, Florida, ed. A. Wu, 20–23.

Bedau, M. 2003. Downward causation and the autonomy of weak emergence. *Principia, Revista Internacional de Epistemologia* 6: 5–50.

Bedau, M. 2009. The evolution of complexity. In *Mapping the future of biology*, eds. A. Barberousse, and T. Pradeu, 211–230. Dordrecht: Springer.

Bedau, M., E. Snyder, and N. Packard. 1998. A classification of long-term evolutionary dynamics. In *Artificial life*, vol. VI, ed. C. Adami et al., 189–198. Cambridge: MIT Press.

Bonner, J.T. 1998. *The evolution of complexity by means of natural selection*. Princeton: Princeton University Press.

Carroll, S. 2005. *Endless forms most beautiful*. Sunderland: Sinauer.

Channon, A., and R. Damper. 1998. Perpetuating evolutionary emergence. In *From animals to animats*, ed. R. Pfeifer, B. Blumberg, J.A. Meyer, and S. Wilson, 534–539. Cambridge: MIT Press.

Channon, A., and R. Damper. 2000. Towards the evolutionary emergence of complex increasingly advantageous behaviours. *International Journal of Systems Science* 31(7): 843–860.

Chu, J., and C. Adami. 1999. A simple explanation for taxon abundance patterns. *PNAS* 96(26): 15017–15019.

Cracraft, J. 2000. The origin of evolutionary novelties: Pattern and process at different hierarchical levels. In *Evolutionary innovations*, ed. M. Nitecki, 21–43. Chicago: University of Chicago Press.

Crubelier, B., P. Preux, and C. Cambier. 1997. Studying adaptation with Echo. ftp://ftp.cogs.susx.ac.uk/pub/ecal97/online/F045.ps.gz.

Crutchfield, J., and M. Mitchell. 1992. The evolution of emergent computation. *Proceedings of the National Academy of Science* 92(23): 10742.

Crutchfield, J., M. Mitchell, and P. Hraber. 1993. Revisiting the edge of chaos: Evolving cellular automata to perform computations. *Complex systems* 7: 89–130.

Damuth, J., and L. Heisler. 1988. Alternative formulations of multilevel selection. *Biology and Philosophy* 4(3): 407–430.

Dawkins, R. 1982. *The extended phenotype*. New-York: Oxford University Press.

Dennett, D. 1995. *Darwin's dangerous idea*. New-York: Simon & Shuster.

Eldredge, N. 1985. *The unfinished synthesis*. New-York: Oxford University Press.

Epstein, J. 1999. Agent-based computational models and generative social science. *Complexity* 4(5): 41–60.

Fontana, W., and L. Buss. 1994. What would be conserved if 'the tape were played twice'? *PNAS* 91: 757–761.

Fontana, W., and P. Schuster. 1998. Continuity in evolution: On the nature of transitions. *Science* 280: 1451–1455.

Forrest, S. 1996. Genetic algorithms. *ACM Computer surveys* 28(1): 77–80.

Forrest, S., and T. Jones. 1994. Modeling complex adaptive systems with Echo. In *Complex systems; mechanisms of adaptation*, ed. R.J. Stonier and X.H. Yu, 3–21. Amsterdam: Ios Press.

Füchslin, R., and J. Mc Caskill. 2001. Evolutionary self organisation of cell-free genetic coding. *PNAS* 98(16): 9185–9190.

Gayon, J. 1998. *Darwinim's struggle for survival*. Cambridge: Cambridge University Press.

Gilbert, S.F., Opitz, G., and Raff, R. 1996. Resynthesizing evolutionary and developmental biology. *Development and evolution* 173: 357–372.

Godwin, B. 1996. *Form and transformation: Generative and relational principles in biology*. Cambridge: Cambridge University Press.

Golberg, D.E., K. Deb, and B. Korb. 1991. Don't worry, be messy. In *Proceedings of the fourth international conference on genetic algorithms*, ed. R.K. Belew and L.B. Booker, 24–30. San Mateo: Morgan Kaufmann.

Gould, S.J. 1977. *Ontogeny and phylogeny*. Belknap: New Haven.

Gould, S.J., and N. Eldredge. 1977. Punctuated equilibria. *Paleobiology* 3: 115.

Gould, S.J., and E. Vrba. 1982. Exaptation: A missing term in the science of form. *Paleobiology* 8: 4–15.

Grüne-Yanoff, Till. 2009. The explanatory potential of artificial societies. *Synthese* 169: 539–555.

Hedrich, R. 1999. The sciences of complexity: a kuhnian revolution in sciences. *Epistemologia* XII.1.

Hillis, D. 1998. *The pattern on the stone: The simple ideas that make computers work*. London: Basic Books.

Holland, J. 1995. *Hidden order. How adaptation builds complexity*. Readings: Helix.

Hubbell, J. 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton: Princeton University Press.

Humphreys, P. 2004. *Extending ourselves*. Oxford: Oxford University Press.

Humphreys, P., and P. Huneman. 2008. Dynamical emergence and computation: An introduction. *Minds & Machines* 18: 425–430.

Huneman, P. 2008a. Combinatorial vs. computational emergence: Emergence made ontological? *Philosophy of Science* 75: 595–607.

Huneman, P. 2008b. Emergence and adaptation. *Minds and Machines* 18: 493–520.

Huneman, P. 2010. Determinism and predictability: lessons from computational emergence. *Synthese*, on-line first.

Kauffmann, S. 1993. *The origins of order*. Oxford: Oxford University Press.

Kettlewell, R. 1956. Further selection experiments on industrial melanism in the Lepidoptera. *Heredity* 10: 287–301.

Kingsolver, J.G., and M.A.R. Koehl. 1985. Aerodynamics, thermoregulation, and the evolution of insect wings: Differential scaling and evolutionary change. *Evolution* 39: 488–504.

Kirschner, M., and G. Gerhardt. 2005. *The plausibility of life: Resolving Darwin's dilemma*. New Haven: Yale University Press.

Koza, J.R. 1992. *Genetic programming: On the programming of computers by means of natural selection*. Cambridge: MIT Press.

Küppers, G., and J. Lenhard. 2005. Validation of simulation: Patterns in the social and natural sciences. *Journal of Artificial Societies and Social Simulation* 8(4). http://jasss.soc.surrey.ac.uk/8/4/3.html.

Lange, M. 1996. Life, 'Artificial life' and scientific explanations. *Philosophy of Science* 63: 225–244.

Langton, C. 1996. Artificial life. In *The philosophy of artificial life*, ed. M. Boden. Oxford: Oxford University Press.

Lenski, R., C. Ofria, T. Pennock, and C. Adami. 2003. The evolutionary origin of complex features. *Nature* 423: 139–144.

Maley, C. 2001. Comparing causal factors in the diversification of species. In *Studies in complex systems*, vol. II, ed. Y. Bar-Yam.

Margolus, D., and J. Toffoli. 1987. *Cellular automata*. Cambridge: MIT Press.

Margulis, L., and D. Sagan. 1986. *Origins of sex: Three billion years of genetic recombination*. New Haven: Yale University Press.

Maron, M. 2004. Evolution of industrial melanism: A spatial, predator-prey genetic algorithm. www.brainoff.com/easy/moth/report.pdf.

Maynard, Smith J., and E. Szathmary. 1995. *The major evolutionary transitions*. New-York: Oxford University Press.

Maynard-Smith, J. 2000. The concept of *information* in biology. *Philosophy of Science* 67: 177–194.

Mayr E. 1976. The emergence of evolutionary novelties. In *Evolution and the diversity of life*, 87–113. Cambridge: Harvard University Press.

Mc Shea, D. 1994. Mechanisms of large-scale evolutionary trends. *Evolution* 48(6): 1747–1763.

Mc Shea, D. 1996. Metazoan complexity and evolution: Is there a trend? *Evolution* 50(2): 477–492.

Mc Shea, D. 2005. The evolution of complexity without natural selection: A possible large-scale trend of the fourth kind. *Paleobiology* 31(2): 146–156.

Michod, R. 1999. *Darwinian dynamics*. Oxford: Oxford University Press.

Michod, R. 2005. On the transfer of fitness from the cell to the multicellular organism. *Biology and Philosophy* 20: 967–987.

Miller, G., and P. Todd. 1995. The role of mate choice in biocomputation; sexual selection as a process of search, optimization and diversification. In *Evolution and biocomputation: Computational models of evolution*, ed. W. Banzhaf and F.H. Eckmann, 169–204. Berlin: Springer.

Mitchell, M. 1998. *An introduction to genetic algorithms (complex adaptive systems)*. Cambridge: MIT Press.

Mitchell, M., and S. Forrest. 1993. Genetic algorithms and artificial life. *Artificial Life* 1(3): 267–289.

Müller, G. 2000. Developmental mechanisms at the origin of evolutionary novelty: A side-effect hypothesis. In *Evolutionary innovations*, ed. M. Nitecki, 99–130. Chicago: University of Chicago Press.

Müller, G.B. 2007. Evo-devo: Extending the evolutionary synthesis. *Nature Reviews Genetics*, published online 6 Nov 2007.

Müller, G.B., and S. Newman. 2005. The innovation triad: An EvoDevo agenda. *The Journal of Experimental Zoology* MDE 304: 487–503.

Müller, G.B., and G. Wagner. 2003. Innovation. In *Keywords and concepts in evolutionary developmental biology*, ed. B.K. Hall and W. Olson, 218–227. Cambridge: Cambridge University Press.

Newman, S.A., and G.B. Müller. 2005. Origination and innovation in the vertebrate limb skeleton: An epigenetic perspective. *The Journal of Experimental Zoology* MDE 304: 593–609.

Nilsson, D.-E., and S. Pelger. 1994. A pessimistic estimate of the time required for an eye to evolve. *Proceedings of the Royal Society of London, Biological Sciences* 256: 53–58.

Okasha, S. 2006. *The levels of selection in evolution*. Cambridge: Cambridge University Press.

Pattee, H. 1989. Simulations, realizations and theories of life. In *Artificial life*, ed. C. Langton et al., 63–78. London: Addison Wesley.

Raff, R. 1996. *The shape of life. Genes, development and the evolution of animal form*. Chicago: University of Chicago Press.

Ray, T. 1992. An approach to the synthesis of life. In *Artificial life*, vol. II, ed. C. Langton, 371–408. London: Addison Wesley.

Reynolds, C. 1987. Flocks, herds and schools: A distributed behavioural model. *Computer graphics* 21(4): 25–34.

Sayama, H. 2003. Spontaneous evolution of self reproducing loops in cellular automata. In *Studies in complex systems*, ed. Bar Yam, 362–377.

Sayama, H., A. Antony, and C. Salzberg. 2003. Genetic diversification and complex genealogy of self-replicators discovered in simple cellular automata: A preliminary report. *Journal of Three Dimensional Images* 17(4): 103–109.

Simpson, G.G. 1949. *Meaning of evolution*. New Haven: Yale University Press.

Sober, E. 1991. Learning from functionalism: prospects for strong artificial life. In *Artificial life II*, ed. C. Langton, C. Taylor, J.D. Farmer, and S. Rasmussen, 749–765. Redwood City, CA: Addison-Wesley.

Wagner, A. 1999. Causality in complex systems. *Biology and Philosophy* 14: 83–101.

Wagner, G., and L. Altenberg. 1996. Complex adaptations and the evolution of evolvability. *Evolution* 50(3): 967–976.

Watson, R. 2005. *Compositional evolution. The impact of sex, symbiosis and modularity on the gradualist framework of evolution*. Cambridge: MIT Press.

Wilke, C., and C. Adami. 2003. Evolution of mutational robustness. *Mutation Research* 522: 3–11.

Williams, G. 1992. *Natural selection: Domains, levels and challenges*. Oxford: Oxford University Press.

Williamson, P.G. 1981. Morphological stasis and developmental constraint: Real problems for neo-darwinism. *Nature* 294: 214–215.

Winsberg, E. 2003. Simulated experiments: Methodology for a virtual world. *Philosophy of Science* 70: 105–125.

Chapter 10
Evolutionary Psychology and the Unity of Sciences: Towards an Evolutionary Epistemology

Luís Moniz Pereira

Abstract This work concerns a non-traditional approach to the unity of sciences, based on a challenging, albeit conjectural, articulation of views proceeding from Evolutionary Psychology and Biology, non-monotonic and decision Logics, and Artificial Intelligence.

The resulting amalgam sets forth a consilience stance, wherefore the unity of science is heuristically presupposed by means of a set of pragmatic and productive default assumptions. It is by virtue of them that we conduct scientific inquiry, the consilience arising from a presumed unity of objective reality, itself of a heuristic and pragmatic conception.

The attending hinges to Artificial Intelligence inevitably suggest the emergence of an innovative symbiotic form of evolutionary epistemology.

Keywords Consilience • Epistemology • Evolutionary Psychology • Logic, Artificial intelligence

10.1 Consilience

In his 1941 classic *Man on His Nature*, the British neurobiologist Charles Sherrington spoke of the brain as an enchanted loom, perpetually weaving a picture of the external world, tearing down and reweaving, inventing other worlds, creating a miniature universe. The communal mind of literate societies – world culture – is an immensely larger loom. Through science the brain has gained the power to map external reality far beyond the reach of a single mind, and through the arts

---

L.M. Pereira (✉)
Centro de Inteligência Artificial – CENTRIA, Departamento de Informática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa – UNL, 2829-516 Caparica, Portugal
e-mail: lmp@di.fct.unl.pt

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology, and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_10,

the means to construct narratives, images, and rhythms immeasurably more diverse than the products of any solitary genius. The loom is the same for both enterprises, for science and for the arts, and there is a general explanation of its origin and nature and thence of the human condition, proceeding from the deep history of genetic evolution to modern culture. Consilience of causal explanation is the means by which the single mind can travel most swiftly and surely from one part of the communal mind to the other.

Arguments in favour of the unity of knowledge – consilience – have been strongly put by Edward O. Wilson, a creator of sociobiology, and author of *Consilience – The Unity of Knowledge* (1988). He postulates there is a single physical nature, and one not persuadable through argumentation or persuasion, whatever the deconstructionists may think. Science is not mere convention.

Consilience, according to him, is the result of co-evolution involving (cultural) memes and genes (see below). Our cultural memes have a genetic basis and cannot, in the long run, stand against the genes who guarantee their survival, although such attempts may potentially exist – viz. through genetic manipulation.

On the other hand, we have several different cultures, though these are produced by brains which have evolved to solve similar problems in ancestral times and, as such, cannot be exceedingly different or distant. Consilience puts in check the romantic conception of the mind as a *tabula rasa* and condones not artistic "irreducibility". In the latter conception there a special something which cannot be reduced or converted to anything else, and, as such, prevents science from addressing the realm of art, even if art itself is the product of a brain which has been evolving for millions of years. Consilience considers scientifically approachable human universals, and that way opens a passageway to that missing link between science and art.

10.2 Evolution and the Brain

The first bipedal primates establish the separation between the human species and the other simians. To fathom the abilities of the human brain it is necessary to understand what exactly were the problems our ancestor primates were trying to solve that led them to develop such an extraordinarily intricate brain. We cannot look at the modern human brain, and its ability to create science, as if the millions of evolution-years which attuned it to its present configuration had never taken place. Among the eventual problems we certainly have those of status, territorialism, mating, gregariousness, altruism versus opportunism, the building of artefacts, and the mappings of the external world.

To the *Homo Sapiens Sapiens*' brain, considered indistinguishable from our current one, we assign an estimated an age of one or two hundred thousand years. The Palaeolithic started at about 60 or 30 thousand years before that, the period in which language, and much later writing, began to develop.

By the Upper Palaeolithic times however, from 40,000 to 10,000 before the present, the tempo of cultural evolution quickened dramatically. According to the theory of population genetics, most of the change was far too fast to be tracked closely by genetic evolution.

As the psychiatrist must look at a patient's past in order to better understand him in the present, so must we look also at our species' past in order to understand our modern peculiarities. This stance is called Evolutionary Psychology – a fascinating field of study – born some 40 years ago.

Evolutionary Psychology is a consummate example of successful ongoing scientific unification, engendered by a deeply significant combination of Psychology, Anthropology, Archaeology, Evolutionary Biology, Linguistics, Neurosciences, and Artificial Intelligence (Buss 2005).

Evolutionary Psychology has been studying the brain from the evolutionary perspective, thereby originating some extremely relevant contributions. In that perspective, it has been strongly supported by Anthropological Archaeology in its empirical study of the cultural evolution of mankind (Stephen 2002).

10.3 Evolutionary Psychology: Genes and Memes

In human life, we have two reproductive mechanisms: one is sexual reproduction, in which the replication unit is the gene; the other is mental reproduction. Some authors from Evolutionary Psychology have construed the notion of "meme", in complement and contrast to the gene. A meme is that which substantiates a second reproductive system executed in the brain. It is the mental unit corresponding to the gene. Memes gather in assemblies, in patterns, similar to the way genes gather in chromosomes. Memes are patterned by ideologies, religions, and common sense ideas. Indeed, certain memes work well together, mutually reinforcing each other, others not, so that correcting (and correctional) mechanisms may be triggered.

We have a genetic reproduction system and, on top of it, Nature – through evolution – has created a second one, which we employ in pedagogy. We reproduce ideas: generally, good ideas propagate and replicate, being selected over the bad ones, although no one is around to guarantee it.

Genes persist because they reproduce, and memes are the reproduction units associated with the brain – the brain being a reproductive organ. What we are doing, in schools and universities, is to reproduce knowledge. Educational systems consist of a means for "infecting" students with good memes, ideas having proven themselves able enough to self-reproduce and persist, while despising others that fail to survive. There are however different variants of educational systems, for instance madrasas.

When they interact, people communicate ideas, and those which are infectiously good tend to reproduce. There are assemblies of ideas, sets of beliefs, which reproduce together. The memes in such memeplexes – like the genes in chromosomes – are in competition amongst themselves and also with the gene base. They exist because they are part of a reproductive mechanism which is necessary to achieve faster local adaptations, as genes take too long to reproduce with respect to the time scale of the individual bearing the memes. Thus the individual phenotype may be given more of a chance to reproduce its genotype. This leads directly to the meme-gene co-evolution.

Memes however could not spread but for the biologically valuable tendency of individuals to imitate, something afforded by the brain. There are plenty of good reasons why imitation should have been favoured by conventional natural selection working on genes. Individuals that are genetically predisposed to imitate enjoy a fast track to skills that may have taken others a long time to build.

Consequently, the brain and its accompanying mind are the result of a deep symbiosis, a genetic product influenced by the mechanism of memetic reproduction. In this faster system of adaptation we have reached the point of being able to predict our own memetic (and genetic) mutations, as necessary changes to prepare for the future by anticipating it. That is why we imagine the future – we create hypothetical scenarios, predict the possible futures, and choose to pursue some of them. This is the basis of the battleground of free will, a useful product of evolution – the ability to imagine scenarios and prefer among them through enacting choices.

However, beyond simple reproductive success there are also pressing concerns in regard to social interaction. As communal beings, we need to develop some sort of status in order to be respected, copied, or obeyed. We must worry about territorial expansion and its defence, if we are to have descendants and, moreover, descendents with descendents. We need to sign contractual agreements with those who share our social and cultural ecology. There is also the important requisite of personal expression opportunity. If we do not express ourselves, no one will copy even our dearest memes, let alone our scientific theory memeplexes.

In this view, scientific thought emerges from distributed personal interaction, albeit it at a special and temporal distance, and never in an isolated way. It must be erected from several confluences, or in teams, as is the case in science. In truth, knowledge is not constructed in an autonomous way; rather it is engendered by networks of people. In science it is important to work as a team. The stereotype of the isolated and enlightened aristocratic scientist has been defeated for quite some time: science is institutionalized, organized and has proper methodologies, conferences. It is processed in appropriate environments, one of them being the educational one, in which we carry out *memetic* proliferation.

Language is the instrument which allows us to fabricate knowledge together, because there is no isolated thought. We go so far as to state that there is no isolated consciousness, that all consciousness is distributed. In particular, any idea of a genius-like isolated consciousness is a myth. When we consider consciousness we should take it out of the brain and spread it through culture, and this is the importance of language.

10.4 Specific Modules Versus General Intelligence

Theoretical and field archaeologists, like Steven Mithen in *The Prehistory of Mind* (1996), are bringing in historical and pre-historical evidence that our ancestors began with a generic intelligence, such as we find in apes.

There has been a broad discussion – in fact reproduced within the Artificial Intelligence (AI) community – about whether intelligence is a general functionality or else best envisaged as divided into specific ability modules or components. When it first appeared, Evolutionary Psychology developed a trend, which Chomsky had begun in insisting on innate specialized areas for language processing in the brain, and it was generally accepted that a plethora of specific modules for a diversity of certain brain functions do exist. Indeed, in the beginnings of Evolutionary Psychology, people like Steven Pinker, Leda Cosmides, John Tooby, and David Buss, in consonance with AI's own vision of specific modules, believed that all brain functions were founded on such modules – there would be modules for language, for mating, religion, etc.

Meanwhile, archaeologists have come to demonstrate, through their historical records, that human species went from a first phase of general intelligence to a second phase of three major specialized modules: one for natural history and naive physics (knowledge of Nature); one for knowledge and manufacture of instruments; and one for cultural artefacts, i.e. the rules of living in society and the very politics of coexistence.

These three specialized intelligences were separated, and it is only at a newer stage – corresponding to *Homo Sapiens*, with the appearance of spoken language – that it becomes necessary to have a cupola module, articulating the other modules. How else do the different specialized modules connect, and how can people communicate amongst themselves? That need gave birth to the generic cupola module, a more sophisticated form of general intelligence, the cognitive glue bringing the specialized modules to communicate and cooperate.

10.5 The Evolution of Reason: Logic

The formal systems of logic have ordinarily been regarded as independent of biology, but recent developments in evolutionary theory suggest that biology and logic may be intimately interrelated. William S. Cooper (2001) outlines a theory of rationality in which logical law emerges as an intrinsic aspect of evolutionary biology.

This biological perspective on logic, though at present unorthodox, could change traditional ideas about the reasoning process. Cooper examines the connections between logic and evolutionary biology and illustrates how logical rules are derived directly from evolutionary principles, and therefore have no independent status of their own. Laws of decision theory, utility theory, induction, and deduction are reinterpreted as natural consequences of evolutionary processes. Cooper's connection of logical law to evolutionary theory ultimately results in a unified foundation for an evolutionary science of reason.

According to Cooper, today, in the general drift of scientific thought, *logic* is treated as though it were a central stillness. For the most part, the laws of logic are taken as fixed and absolute. Contemporary theories of scientific methodology are logico-centric. Logic is seen commonly as an immutable, universal, meta-scientific framework for the sciences, as for personal knowledge. Biological evolution is acknowledged, but it is accorded only an ancillary role, as a sort of biospheric police force, whose duty is to enforce the logical law among the recalcitrant. Logical obedience is rewarded and disobedience punished by natural selection, it is thought. All organisms with cognitive ability had better comply with the universal laws of logic on pain of being selected against!

Comfortable as that mindset may be, Cooper believes he is not alone in suspecting it has things backward. There is a different, more biocentric, perspective to be considered. In the alternative scheme of things, logic is not the central stillness. The principles of reasoning are neither fixed, absolute, independent, nor elemental. If anything, it is the evolutionary dynamic itself that is elemental. Evolution is not the law enforcer but the law giver – not so much a police force but a legislature. The laws of logic are not independent of biology but implicit in the very evolutionary processes that enforce them. The processes determine the laws.

If the latter understanding is correct, logical rules have no separate status of their own but are theoretical constructs of evolutionary biology. Logical theory ought then in some sense to be deducible entirely from biological considerations. To paraphrase, the hypothesis is that the commonly accepted systems of logic are branches of evolutionary biology.

Indeed, evolution has provided humans with symbolic thought, and symbolic language communication abilities. Objective common knowledge requires thought to follow abstract, content independent rules of reasoning and argumentation, which must not be entirely subjective, on pain of making precise communication and collective rational endeavour impossible. Such rules have become ingrained in human thought, and hold an enormous joint survival value.

In human cognitive evolution, both mimetic knowledge (such as that inherent in reality-simulating maps and models), and imitation knowledge (such as that present in ritual observation, or in artefact reproduction), were essential first steps towards socially situated, joint rule following behaviour, required by, say, hunting plans.

Decision theory is the branch of logic that comes into most immediate contact with the concerns of evolutionary biology. They are bound together by virtue of their mutual involvement in behaviour. The logic of decision is concerned with choices regarding the most reasonable courses of action, or behavioural patterns. Behaviour is observable, it is amenable to scientific prediction and explanation, and there is the possibility of explaining it in evolutionary terms. This makes behaviour an interdisciplinary bridge approachable from both the biological and the logical sides. Ultimately, behaviour is the fulcrum over which evolutionary forces extend their leverage into the realm of logic. Viewed through the lenses of biology, favoured behaviour is evolutionary fit. Through the lens of logic it is rational decision behaviour (Cooper 2001), according to rules for reasoning and rules for action.

On the heels of rational group behaviour, throughout human cultures there emerged abstract rule following social games. Game rules encapsulate concrete situation defining patterns, and concrete situation-action-situation causal sequencing, which mirrors causality-obeying physical reality. From games, further abstraction ensued, and there finally emerged the notions of situation-defining concepts, of general rules of thought and their chaining, and of legitimate argument and counter-argument moves. Together they compose a cognitive meta-game (Holland 1998).

The pervasiveness of informal logic for capturing knowledge and for reasoning, a veritable *lingua franca* across human languages and cultures rests on its ability to actually foster rational understanding and common objectivity. Crucially, objective knowledge evolution dynamics, whether individual or plural, follows ratiocination patterns and laws.

Furthermore, and more recently, the very same rules of reasoning can and are employed to reason about reasoning. Moreover, new reasoning methods can and have been invented and perfected throughout human history. Examples of these are transfinite induction, *reductio ad absurdum* (proof by contradiction), recursion, abduction, and contradiction removal, to name but a few.

Though some reasoning methods are well known, some are still unconscious but, like the rules of grammar, can be discovered through research. Indeed, humans can use language without learning grammar. However, if we are to understand linguistics, knowing the logic of grammar, syntax and semantics is vital. Humans do use grammar without any explicit knowledge of it, but that doesn't mean it cannot be described. Similarly, when talking about the movement of electrons we surely do not mean that a particular electron knows the laws it follows, but we are certainly using symbolic language to describe the process, and it is even possible to use that description to implement a model and simulation which exhibits precisely the same behaviour.

New purported reasoning methods may be disputed, just like any specific train of reasoning can. But reasoning can only be disputed by further reasoning, if any consensus is to be found! (Nagel 1997). Some argue that scientific and philosophical discussion is necessarily a tantamount to a culture sensitive, and culturally relative, persuasive informal *ad hoc* argumentation, allied to anything goes rhetoric (criticized by Paul Gross and Norman Levitt 1994). They ignore that argumentation is just another form of reasoning which has itself been made the subject of logical formalization, and are oblivious to the fact that rhetoric may be fine for preachers, but is not conducive to the two-sided communication required to reach common agreement in the all rigorous scientific praxis that lead to cumulative knowledge.

Logic, we sustain, provides the overall conceptual cupola that, as a generic module, fluidly articulates together the specific modules identified by evolutionary psychology. In that respect, it is mirrored by the computational universality of computing machines, which can execute any program, compute any computable function.

The relationship of this argument to logic is ensured by the philosophical perspective of functionalism: logic itself can be implemented on top of a symbol processing system, independently of the particular physical substrate supporting it. Once a process is described in logic, we can use the description to synthesize an artefact with those very same properties. As long as it is a computational model, any attempt to escape logic will not prove itself to be inherently more powerful.

On the other hand, there is an obvious human capacity for understanding logical reasoning, a capacity developed during the course of brain evolution. Its most powerful expression today is science itself, and the knowledge amassed from numerous disciplines, each of which with their own logic nuances dedicated to reasoning within their domain. From nation state laws to quantum physics, logic, in its general sense, has become the pillar on which human knowledge is built and improved, the ultimate reward for our mastery of language.

10.6 Realism and the Unity of Sciences: Our Stance

Belief in the intrinsic unity of knowledge, whatever may be its reliance on logic, rides ultimately on the hypothesis that every mental process has a physical grounding and is consistent with the natural sciences. The mind is supremely important to the consilience program for a reason both elementary and disturbingly profound. Everything that we know and can ever know about existence is created or absorbed there.

We partake of a species which evolved a brain that copes with its doubly situated existence in nature and nurture. And in this endeavour it is enabled by the wherewithal in jointly modelling and changing both one and the other. The universal plasticity and the mimetic ability of the human mind account for its success in striving for and achieving consilience.

Nevertheless, all that has been learned empirically about evolution in general, and mental processes in particular, suggests that the brain is a machine assembled not to understand itself, but to survive. Understanding the mind at work, then, needs to be brought about by the methods of science.

The human attainment of high intelligence and culture ranks as the last four great steps in the grand history of life. They followed one upon the other at roughly one-billion-year intervals. The first was the beginning of life itself, in the form of simple bacterium like organisms. Then came the origin of the complex eukaryotic cell through the assembly of the nucleus and other membrane-enclosed organelles into a tightly organized unit. With the eukaryotic building block available, the next advance was the origin of large, multicellular animals such as crustaceans and molluscs, whose movements were guided by sense organs and central nervous systems. Finally, there came humanity and its cortex, with the ability to perform science and to change the world.

However, what we know of the heredity and development of the brain shows them to be almost unimaginably complicated. The human genome database reveals it to be comprised of over 30,000 genes, with at least over 3,000 distinct ones. The molecular processes that guide the growth of neurons to their assigned places have only begun to be deciphered. Overall, the human brain is the most complex object known in the universe.

Notwithstanding, here is this paper's stance on the Unity of Sciences:

- At some point, it seems a materialist pragmatic heuristic to believe, i.e. to introduce a default postulate, to the effect that a unifying consilience of mind and body will be met.
- Furthermore, we are entitled to pragmatically and heuristically presuppose that the brains we have in common, received via ancestral evolution, are indeed

capable of ever extendable joint agreement regarding the scientific view of our shared reality, especially in view of our brains' plasticity of communication and modelling.
- Finally, we can pragmatically, and for efficiency's sake, assume that the very unity of mind-independent reality (a presumed given) is thereby conducive to the unity of the sciences themselves.

These productive and tenable working assumptions have yet to be disproved (even in spite of postmodernism . . .) and so we keep to them. Let us dub the position that goes with them "Evolutionary Pragmatic Epistemological Realism", inspired by Nicholas Rescher's *Realism and Pragmatic Epistemology* (2005).

And we presume thence a mind-independent reality for at least six important reasons:

- To preserve the distinction between true and false with respect to factual matters and to operate the idea of truth as agreement with reality.
- To preserve the distinction between appearance and reality, between our *picture* of reality and reality itself.
- To serve as a basis for intersubjective communication.
- To furnish the basis for a shared project of communal inquiry.
- To provide for the fallibilistic view of human knowledge.
- To sustain the causal mode of learning and inquiry and to serve as a basis for objectivity of experience.

What is at stake in the present stance is ultimately a principle of practice, and thought practice to be sure. Accordingly, the justification for our fundamental presuppositions is not *evidential* at all; postulates as such are not based on evidence. Rather, it is practical and instrumentalistic – pragmatic, in short. It is procedural or functional efficacy that is the crux. The justification of these postulates lies in their utility: we need them to operate our conceptual scheme. Consequently, our unity of science stance's epistemic status is not that of an empirical discovery but of an encompassing presupposition whose ultimate justification is a transcendental argument from the very possibility of communication and inquiry as we typically conduct them.

10.7 Postmodernism

Now turn we to postmodernism, the ultimate polar antithesis of the Enlightment. The difference between the two extremes can be expressed roughly as follows: Enlightment thinkers believe we can know everything, and radical postmodernists believe we can know nothing. The philosophical postmodernists challenge the very foundations of science and traditional philosophy. Reality, they propose, is a state constructed by the mind, not perceived by it. In the most extravagant version of this constructivism, there is no "real" reality, no objective truths external to mental activity, only prevailing versions disseminated by ruling social groups.

Postmodernism is expressed more explicitly still in “deconstruction”, a technique of literary criticism. Each author’s meaning is unique to him, goes the underlying premise; nothing of his true intention, or anything else connected to objective reality, can be reliably assigned to it. His text is therefore open to fresh analysis and commentary issuing from the equally solipsistic head of the reviewer. But then the reviewer in turn is subject to deconstruction.

Patently, postmodernism puts itself into question in inescapable self paradox as a method to obtain secure knowledge, and is incompatible with scientific methodology. The latter relies on the existence of a regularity abiding external reality, which cannot be emotionally cajoled, and which is both human history and society independent.

10.8 Epistemic Tools

However, the canonical definition of objective scientific knowledge avidly sought by the logical positivists is not a philosophical problem nor can it be attained, as they hoped, simply by logical and semantical analysis. It is an empirical question too, that can be answered only by a continuing probe of the possible functionality of the thought process itself and its physical basis.

In some cases, the cognitive tools and instruments of rationality will be found hardware independent. Even then, the appropriateness of their use in specific real circumstances and goals will need to be empirically determined. There is no universal one-size-fits-all epistemological recipe, but agreement can be had on the relative success of any given tool kit.

In any case, partial understanding may also be sought by building intelligent machines, functionalism coming to the rescue when positing that the material substrate is often not of the essence, that it suffices to realize equivalent functionality albeit over different hardware. Moreover, distinct functioning roads to the same behaviour may be had, thereby accruing to our understanding of what general intelligence means, toward their symbiotic entwining, the most recent step in evolutionary epistemology. Functionalism can only make that more adroit.

The most fruitful procedures will almost certainly include the use of Artificial Intelligence, theory and technique, aided in time by the still embryonic field of artificial emotion, to simulate complex mental operations. This modelling system will be joined to an already swiftly maturing neurobiology of the brain, including the high-resolution scanning of computational networks active in various forms of thought. Important advances are also deemed to come eventually from the molecular biology of the learning process.

How does natural selection anticipate our future needs? Well, by creating a cognitive machine called brain that can create models of the world, and even of itself, and process hypotheticals much like a Universal Turing Machine can mimic any other Turing machine, and just like any given computer can run any program. This plasticity provides for its universal versatility (cf. Davis 2000).

It is useful to consider a duality I designate "Turing versus Eve". The mathematician Alan Turing represents the computer in the essence of its complete flexibility. The Universal Turing Machine is the one which can imitate every computer, it is mimetism *par excellence.* That mimetism makes us think about the meme and our own mental flexibility, so vital in complementing genetic reproduction, due to the different reproduction timings. In the latter, the difference spans across generations, and that is not enough when adaptation must be agile. It is from that need that stems the cerebral mechanism of reproduction – those memes which jump from brain to brain. In genetic reproduction, mitochondria are genetic structures from the feminine side which are replicated without mating of genes. They correspond to the specific modules we inherit in virtue of our species' past.

With this background in mind, and namely the discussion about specialized modules and general intelligence, I would like introduce at this point the informal notion of *cognome*, by analogy with genome, standing for an individual's particular structural combination of cognitive memes.

When considering scientific knowledge, if the computer processing of the human genome is what leads us to Bio-informatics then, by analogy, we may state that the cognome will be the basis of a future "Cognotechnology", applicable in any science. This way, the future of AI is connected to the characteristic of it being an epistemological instrument, not only for an autonomous agent, but a symbiotic one which will help humans in performing science itself.

And I'm not just talking about data mining, pattern recognition, ontology building, although in those fields we can approach more structured aspects of epistemology. I'm thinking about that which every scientist does, which is to abduce, invent and prophesy theories, put them to the test, create experiments, draw conclusions to support additional observations, discuss those observations and his conjectures with other scientists.

There is an ongoing meta-argumentation about what is good reasoning, what are the conclusions we can draw from a discussion (i.e. a semantics), which is inherent to all scientific activity. The computer will be used more and more as a research aide, not just to automate but also propose experiences and hypotheses and, in the end, by making our own conceptions on epistemology application repeatable and externalized, it will make them more objective too.

Veritably, the capacity for cognition is what allows us to anticipate the future, to pre-adapt and imagine scenarios of possible evolutions – of the world and of ourselves as cognitive agents – to make choices, to use preferences about some hypothetical worlds and their futures, and meta-preferences – preferences on which preferences to employ and how to make them evolve.

The activity of prospecting the future is vital and characteristic of our species and its capacity to understand the real world and ourselves, living in society, where distributed cognition is the normal and regular way to do science.

Prospective consciousness allows us to pre adapt to what will happen. For that, a capacity to simulate, to imagine "what would happen if", i.e. is hypothetical thinking, becomes necessary. Such thinking is indispensable in science; for it gives us the rules to predict and explain what will or can happen, without which technology would not be possible.

Lately, I've been working towards automating this capacity, by implementing programs which can imagine their futures, making informed choices about them, and then modify themselves to enact those choices – the inklings of free will. We call it prospective computing (Luís Moniz Pereira and Gonçalo Lopes 2009).

Epistemology will eventually have the ability to be shared, be it with robots, aliens or any other entity who must needs perform cognition to go on existing and program its future. Creating situated computers and robots means carrying out our own cognitive evolution by new means. With the virtue of engendering symbiotic, co-evolving, and self-accelerating loops. Computerized robots reify our scientific theories, making them objective, repeatable, and part of a commonly constructed extended reality, built upon multi-disciplinary unified science.

Artificial Intelligence and the Cognitive Sciences, by building such entities, provide a huge and stimulating step towards furthering Science Unity, through the very effort of that construction. To this end, the functionalist stance is most helpful.

10.9 Coda

Evolution, including genetic progress in human nature and human capacity, will be from now on increasingly the domain of science and technology, tempered by ethics and political choice.

With rare exceptions, universities have dissolved their curriculum into slurries of minor disciplines and specialized courses. A balanced perspective cannot be acquired by studying disciplines in pieces, but through the pursuit of consilience among them. Only fluency across the boundaries will provide a clear view of the world as it really is, not as seen through the lenses of ideologies and religious dogmas, or commanded by myopic response to immediate need.

Moreover, interdisciplinary high level research and communication channels need to be institutionalized, such as in Institutes of Advanced Study.

Last but not least, according to Edward O. Wilson (1998), gene-cultural evolution is the underlying process by which the brain evolved and the arts originated. It is the conceivable means most consistent with the joint findings of the brain sciences, psychology, and evolutionary biology. Still, *direct* evidence with reference to the arts is slender. It is possible that new discoveries concerning the brain and evolution will yet change the picture fundamentally. The uncertainty makes the search for the alignment of science and the humanities all the more interesting a prospect.

References

Buss, David M. (ed.). 2005. *The handbook of evolutionary psychology*. Hoboken: John Wiley & Sons.

Cooper, William S. 2001. *The evolution of reason: Logic as a branch of biology*, Cambridge Studies in Philosophy & Biology. Cambridge: Cambridge University Press.

Davis, Martin. 2000. *The universal computer: The road from Leibniz to Turing*. New York: W.W Norton & Co.

Gross, Paul R., and Norman Levitt. 1994. *Higher superstition*. Baltimore: The Johns Hopkins University Press.

Holland, John. 1998. *Emergence – From Chaos to Order*. Reading: Addison-Wesley.

Pereira, Luís Moniz, and Lopes, Gonçalo. 2009. *Prospective Logic Agents,* International Journal of Reasoning-based Intelligent Systems (IJRIS), 1(3/4):200–208, 2009.

Mithen, Steven. 1996. *The prehistory of mind*. London: Thames and Hudson Ltd.

Nagel, Thomas. 1997. *The last word*. New York: Oxford University Press.

Rescher, Nicholas. 2005. *Realism and pragmatic epistemology*. Pittsburgh: University of Pittsburgh Press.

Shennan, Stephen. 2002. *Genes, memes and human history – Darwinian archaeology and cultural evolution*. London: Thames & Hudson Ltd.

Sherrington, Charles. 1941. *Man on his nature*, The Gifford Lectures, Edinburgh 1937–8. New York: Macmillan.

Wilson, Edward O. 1998. *Consilience – The unity of knowledge*. New York: Alfred A. Knopf.

Chapter 11
Unity of Science and Pluralism: Cognitive Neurosciences of Racial Prejudice as a Case Study

Luc Faucher

Abstract A consideration of the recent history of philosophy reveals that when thinking about unity of science, philosophers have mainly been thinking of unity through reduction of higher level theories to lower level theories. In other words, if unity was to be achieved it was through *intertheoretic reduction*. Lately, though, some philosophers (Darden and Maull, Philos Sci 44:43–64, 1977; McCauley and Bechtel, Theory Psychol 11:736–760, 2001; Mitchell and Dietrich, Am Nat 168:S73–S79, 2006) have started to question this exclusive focus on intertheoretic reduction in the discussions concerning the unity of science. These philosophers have also come to reject the global project of unification for more modest and local forms of unification: This is the area of pluralism. Pluralism complicates tremendously our understanding of the relations between theories. In fact, pluralism suggests that we are facing two distinct tasks: (1) Developing a typology of the intertheoretic relations; (2) Understanding on a case-by-case basis the relation between specific theories or specific frameworks. I believe that progress has been made with respect to (1), but I want to improve on the current understanding of the typology of intertheoretic relations. I take (2) to be essential: Many scientists have failed to understand what pluralism entails. They view their theories to be simply inconsistent with each other (when sometimes, they are not). It is important to understand the relations between actual sciences and between actual theories in order to avoid futile arguments and to develop better theories. In this paper, I will present Sandra Mitchell's typology of inter-theoretic relations. I will then focus on a case study—the relations between the neurosciences and social cognitive psychology of racial prejudice. What will emerge is that the pluralism proposed by Mitchell should be enriched further to understand the real nature of the unity proposed in certain fields of science.

L. Faucher (✉)
Département de Philosophie, Université du Québec à Montréal, CP 8888,
succursale Centre-Ville, Montréal, QC, H3C 3P8, Canada
e-mail: luc.faucher@uqam.ca

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology, and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_11,

Keywords Cognitive neuroscience • Pluralism • Reduction • Racial • Social psychology • Unification

A consideration of the recent history of philosophy reveals that when thinking about unity of science, philosophers have mainly (but not exclusively) been thinking of unity through reduction of higher-level theories to lower level theories (Nagel 1961; Oppenheim and Putnam 1958). In other words, if unity was to be achieved it was through *intertheoretic reduction*. In the optimistic version of intertheoretic reduction, all scientific disciplines would reduce to physics, which would then bear the whole epistemic burden of explanation for all other sciences (this is what is sometimes call "global unification"; Wylie 1999). In its pessimistic version (for instance, Fodor 1974), reduction from higher-level theories to lower-level theories was thought to be impossible and therefore science was condemned to disunity. Lately, though, some philosophers (see Darden and Maull 1977; McCauley and Bechtel 2001; Mitchell and Dietrich 2006) have started to question this exclusive focus on intertheoretic reduction in the discussions concerning the unity of science. These philosophers have also come to reject the global project of unification for more modest and local forms of unification: This is the area of pluralism (Sect. 11.1). Pluralism complicates tremendously our understanding of the relations between theories.

In fact, I think that anyone trying to understand pluralism is facing at least two distinct tasks:

1. Developing a typology of the intertheoretic relations;
2. Understanding on a case-by-case basis the relation between specific theories or specific frameworks

I believe that progress has been made with respect to task 1, and I want to present what I consider as a very interesting step toward a more adequate typology of intertheoretic relations (Sect. 11.2). I take task 2 to be essential: Some scientists have failed to understand what pluralism entails. They view their theories to be simply inconsistent with each other, when indeed sometimes they are not. It is important to understand the relations between actual sciences and between actual theories in order to avoid futile arguments and to develop better theories.

In this paper, I will present Sandra Mitchell's suggestion of a detailed typology of inter-theoretic relations and we will illustrate its utility by focusing on a case study—the relations between the neurosciences and social psychology of racial prejudice (Sects. 11.3 and 11.4). What will emerge is that the Mitchell's model presented in Sect. 11.2 should be enriched further to understand the real nature of the unity proposed (and sought after) in certain fields of science. I will use Todd Grantham's (2004) model of "unity-as-interconnection"—a model descriptively richer and more inclusive than the classical reductive model or of the new reductive model proposed by Bickle 1998, 2003—to show what kind of integration is really achieved in cognitive neurosciences of racial prejudice.

11.1 The End of the Unification Project as We Knew It

In their famous paper, Oppenheim and Putnam (1958) used the expression "unity of science" to describe two things: (1) the ideal state of science (a unitary science) and (2) a trend toward this ideal state. Most philosophers have been following Nagel (1961)[1] in characterizing (1) in terms of a reduction of the theories of the higher-level disciplines to the theories of a lower-level discipline—in principle, physics—where the reduction is obtained when one can derive the higher-level theories from the lower-level, more fundamental, theories.[2]

There is little enthusiasm for the two components of Oppenheim and Putnam's view about the unity of science nowadays. Indeed, in a paper on the unification of science, Philip Kitcher made the following comment concerning the Nagelian version of (1) which, we think, summarizes the recent attitude of most philosophers on the question:

> *The Unity of Science Movement is dead.* If philosophers ever believed that science could be organized as a hierarchy of theories founded on general principles with the basic generalizations of 'higher level' theories derivable from those more 'fundamental' theories, then they do no more. (1999, p. 337; our emphasis)[3]

As for the trend towards unification, as Fodor put it a few years ago:

> It's attending to how the scientific edifice is actually organized that makes the eventual reduction of the rest of science to physics seem so unlikely. Here, for once, 'don't think, look' sounds like a good advice … For what one sees when on looks doesn't at all suggest a structure that is collapsing into its basement. If the unity of sciences is true, then there ought to be fewer sciences every day, as basic physics absorbs them one by one. But what's going on seems to be quite the reverse: an accelerating proliferation of new disciplines: the damn things multiply faster than college deans can keep up with them. (1998; see also Hacking 1996, p. 57)

If one accepts the death sentence of the unification project as carried out by Nagel and his heirs, one seems to have two options: either to re-vamp the project by injecting some new conceptual blood into it or to bury it altogether (e.g., Dupré 1993, 1996). But, as historians of philosophy argued (Creath 1996; Cat, Cartwright and Chang 1996; Richardson 2006), since there is not only one "unity of science" program rejecting one version of the program does not commit one to rejecting "unity of science" in another guise. Indeed, there is a surge of interest for the history of the "unity of science movement" (for a comprehensive history, see Cat 2007) which reveals that even in the heydays of logical positivism there were alternative ways to think about unity that did not imply reduction. For instance, according to Creath (1996), "Neurath never speaks of reducing one science to another but talks

---

[1] "Since Nagel's influential model of reduction by derivation, most discussions of unity of science have been cast in terms of reductions between concepts and between theories" (Cat 1998, p. 532; see also Sarkar 1992, p. 168 for a similar view).

[2] For some clarification on the notion of level, see Sect. 11.2.

[3] Similarly, Mitchell et al. (1997), concludes that the Nagelian project "has failed" (p. 108). Actual causes and time of death will not be discussed here.

instead of symmetrical relations such as connecting, building bridges between, and filling gaps between various branches of science" (161).[4] As it will get clear, my position could be seen as a revival of this other, more liberal, version of the program of unification. To make my position clear, I will use as a foil John Bickle's project of re-vamping the "unity-as-reduction" model.

Recently, John Bickle (2003, 2006) has argued in favor of a general philosophical stance that he calls "new wave metascience." New wave metascience[5] is a bottom-up philosophy of science that tries to capture the sense of reduction as it emerges from actual sciences, "independent of any pre-theoretic or 'philosophical' account of explanation" (2003, p. 31). As he puts it, "the job of new wave metascience is simply to illuminate concepts like reduction as these imbue actual scientific practice" (2003, p. 32). Looking at what he considers the best examples of scientific work on the brain, he observes that 'metascientific reduction' is indeed 'ruthlessly reductionist'. Scientists are not simply reducing the mind to some high-level entities studied by cognitive neurosciences, but they "have developed experimental practices that bridge the behavioral to the molecular pathways levels directly" (2006, p. 414). Contrary to the classical model of reduction, as proposed for instance by Oppenheim and Putnam (1958), in which reduction had to proceed from one level to the next without skipping any level, in actual science, psychological phenomena are explained by going straight down a much lower level (e.g., the molecular), skipping the intermediary levels in between (2006, p. 426). So what science shows, Bickle claims, is, for instance, that it is possible to explain all the known properties of memory in molecular terms, without having to bother explaining them in terms of information processing. He claims that at best, the intermediary levels should be considered as heuristics for discovering explanations at lower levels and that once these explanations have been found, there will nothing left to be explained by the intermediary levels.[6] In that sense, molecular explanations will be fundamental: all the 'real' explanatory work will be done by the lower-level theory (which is now the molecular one, until biophysics takes over), there is nothing (worthwhile) left to be explained by other higher-level theories (higher-level theory are thus 'explanatorily inert'). So while Bickle claims that scientists have abandoned the idea that we should not skip levels to produce a reduction, they have kept alive the epistemic and metaphysic fundamentalism of Oppenheim and Putnam: one theory at a very low level will explain all the higher-level phenomena. The ideal of science is reduction and the trend is toward reduction.

---

[4] For example, Neurath wrote the following: "The development of physicalist sociology does not mean the transfer of laws of physics to living things and their groups ... Comprehensive sociological laws can be found, ..., without the need to go back to the microstructure, and thereby to build up these sociological laws from physical ones" (quoted by Cat et al. 1996, p. 347).

[5] His stance is "new wave" because its attempt to capture some important aspects of the practice internal to science while downplaying the importance of general external philosophical questions can be seen as a reincarnation of the carnapian distinction between 'internal' and 'external' questions (on this, see section 5.2 of Chap.1 of his 2003).

[6] "There is no need to evoke psychological causal explanations, and in fact scientists stop evoking and developing them, once real neurolobiological explanations are on offer" (Bickle 2003, p. 110).

So reductionism "is alive and thriving, at the core of the very core of one of the hottest (and best funded) scientific disciplines" (2003, p. 5), but it is not Nagel's form of reduction. For instance, Bickle considers that a reduction has been achieved in the molecular disciplines when we see how a cellular/molecular story "arranges its constituents together into a sequential and combinatorial structure abstractly similar to psychology's coarse-grained functional posits" (2003, p. 100). The relation between theories is not derivation, but something more akin to mirroring (in his previous book Bickle (1998) considered that the relation between theories is a relation of analogy[7]) where

> . . . entities characterized on the reduced theory primarily by their functional (input–output) features get linked to complex structures (sequences and combinations of entities and processes) whose dynamics and interaction are specifiable within the reducing theory
>
> 1. apply to roughly the same intended set of real-world-systems, and
> 2. provide causal mechanisms that explain the former's functional (input–output) profile (2003, p. 99)

According to Bickle, the explanation in neurosciences is closer to what has been called "mechanistic explanation" (Machamer, Darden and Craver 2000) than to explanation by laws dear to the logical positivists. A mechanistic explanation works by providing a mechanical model or schema of the (molar) behavior of the mechanism. Such a model includes: "(1) a description of the mechanism's behavior; and (2) a description of the mechanism which accounts for that behavior" (Glennan 2002, p. 347), the latter being understood as describing the "guts of the mechanism". But Bickle departs from most recent mechanists in rejecting the idea of multi-level mechanisms. Indeed, most mechanists argue that phenomena are (and should be) explained by a nested hierarchy of models of mechanisms (for an example of such nested hierarchy, see Craver and Darden 2001).[8] By contrast, Bickle thinks that, in certain cases of actual science, the molecular mechanisms are 'tied directly' to the psychological phenomenon and that there is no need for intermediary mechanisms.

In what follows, I want to adopt the "new wave metascience" stance too (I think like Bickle that whether and how the science is unified is an empirical matter, not something to be decided on an a priori grounds or stipulated). I think that, even granting that Bickle is right about memory (which is generous, see for instance Looren de Jong 2006), and granting that the syntactic picture of theory developed by the logical positivists is not the kind of explanatory structure dominant in certain areas of science (that is, explanations in real science are actually mechanistic ones),

---

[7] As he puts it: " . . . the new wave approach construes the relation as the construction of an image . . . of the set-theoretic structure of models of the reduced theory $\mathbf{T_R}$ within the set comprising reducing theory $\mathbf{T_B}$. " (2003, p. 27).

[8] Bechtel and Hamilton (2007) and Faucher (2006) have insisted that while mechanism is typically reductionnist in spirit, it does not have to maintain that a complete causal story can be told at the lower level. Explanations of many phenomena will typically need to make reference to causal interactions of the mechanism with the environment (which will change the conditions of the parts of the mechanism and their mode of operation).

his picture of the relation between theories does not correctly describe a large chunk of what is going on at the intersection of psychology and the neurosciences. As I will show, reduction does not (and will not, as far as I can foresee) accurately describe what is taking place in fields such as the cognitive neurosciences of racial prejudice (Sect. 11.4).[9] In these fields, scientists are not after the global unification advocated both by the logical positivists and by Bickle. However, this does not condemn science to disunity, for as we will see, scientists are concerned with more modest forms of unification that admit a certain degree of plurality. I am convinced that the cognitive neurosciences of racial prejudice are representative of some legitimate scientific work done at the intersection of psychology and neurosciences. As I will show, it cannot be described as a reductive enterprise. If I am right, keeping in mind that Bickle's description of some work in molecular sciences might also be right, we will need a much richer model of unification than the ones proposed until now (which are either reductive or non-reductive). But before attending to this, I would like to further my reflection on unification in order to be more specific on the forms unification can take.

11.2 Pluralism

11.2.1 Competitive Pluralism vs. Compatible Pluralism

As explained in Sect. 11.1, the Unity-of-Science project falters because of its commitment to reduction as derivation as the key intertheoretic relation across levels (and also, to a certain extent, because the observable trend was not toward a unification of the kind advocated). But that's not all: There are additional problems for the Unity-of-Science project. Proponents of this project were also typically committed to some form of *theoretical monism*: They assumed that typically, a single explanatory theory (which, consistent with their reductionism, should be derivable from some lower-level theory) explained the phenomena in a given domain. But, as I shall argue in this section, theoretical monism is only one way to produce unity and it might well be the exception rather than the rule. Let's also say, right off the bat, that I am concerned with epistemic pluralism and not with metaphysical pluralism. I am concerned with whether (and when) there can justifiably be several explanatory theories in a given domain and not with whether there is one world (*metaphysical monism*) or a plurality of worlds (*metaphysical pluralism*) described by these theories.

Theoretical Pluralism—typically, the view that several distinct explanatory theories are explaining the phenomena in a given domain—stands in contrast to theoretical monism. It is useful to follow Mitchell (2002, 2003) in distinguishing

---

[9] Which is not to say that it will not sometimes be an accurate description, see for instance Bickle's remarks about the molecular neurosciences of social recognition memory (2006, p. 423).

two main forms of theoretical pluralism: *competitive* and *compatible* pluralism. The expression "*competitive pluralism*" applies to a domain of phenomena where several *incompatible* explanatory theories are in competition to explain the phenomena in this domain. Theories are incompatible when the correctness of the explanations provided by one of the competing theories entails that the explanations provided by alternative theories are incorrect. Another form of incompatibility, common in the early stages of a discipline, has less often been recognized: Often, theories compete to define what the very phenomena in a given domain are. They disagree about what needs to be explained and what needs not. For instance, relying on the competence/performance distinction (e.g., Chomsky 1965), generative linguists have argued that linguistics is not concerned with the structure of sentences found in corpus, but rather with intuitions of grammaticality.

Philosophers have long recognized competitive pluralism and have used this notion to describe two things: the state of an immature science and the state of a mature science. Kuhn (1962) has argued that immature scientific domains are often characterized by a plurality of theories competing for hegemony. Ultimately, according to Kuhn at least, the plurality should disappear as one theory imposes itself as the more apt to resolve certain problems. Competitive pluralism also describes the state of numerous mature sciences: in cognitive sciences, for instance, two theories were in competition to explain mental imagery, the quasi-pictorial one (Kosslyn 1980) and the descriptional one (Pylyshyn 1981). Philosophers (and scientists) have typically held that competitive pluralism in mature sciences is temporary: It should eventually give way, in what might be a distant future, to unity, that is, one of the theory *has* to be right. As Mitchell puts it:

> The ultimate aim of science is to resolve the conflicts by adopting the single unfalsified, or true, or overwhelmingly supported winner of the competition. (Mitchell 2002, p. 56)

The expression "*compatible pluralism*" applies to a domain of phenomena when several *compatible* explanatory theories explain (or are required to explain) the phenomena in this domain. Mitchell highlights two situations resulting in some form of compatible pluralism. First, there are different kinds of explanation. For instance, as biologists and philosophers have often argued (Tinbergen 1963; Beatty 1994), the same biological phenomena can be explained in several different, but compatible, ways. Following Tinbergen, one might ask what the mechanism underlying a trait is, what its ontogeny is, what its phylogeny is and what its adaptive function is. In cases like these, theories that provide different kinds of explanation are not in competition, because they do not attempt to answer the same question about a phenomenon. Second, as described in detail by Mitchell et al. (1997), a given phenomenon might be the result of several alternative causes. Because different causes produce the division of labor in different species of social insects, the theory that explains the division in labor in bees does not explain the division in labor in ants. More generally, she contends that "it is the diversity of the 'solutions' to adaptive problems and the historical contingencies influencing those variable paths that preclude global, theoretical unification" (2002, p. 67).[10] Moreover, inside a single phylum division of labor can be the product of different factors interacting. To take another example, several causal factors (kinship, punishment, etc.) explain the sociality of many species of social insects (Queller and Strassmann 1998). Explaining the relevant phenomena in such cases will typically proceed by isolating a factor in a model (it will have a *ceteris paribus* clause) that partially explains the phenomenon. Since different models isolate different contributing factors of a single phenomenon, they are not to be understood as being in competition, but as providing a partial explanation of this phenomenon. Because these partial explanations are taken to be explanations of a single actual concrete phenomenon, they might be usefully integrated, if this is possible. However, note that it might not always be possible to integrate these explanations. For example, this might happen when the resulting explanation would be so complex as to be extremely difficult to understand or when approaches of the same phenomena have incongruous 'causal space' (that is, as Longino puts it, when "some phenomena regarded as causally active in one approach are simply not included in the other" (2006, p. 118)).[11] So a further distinction should be introduce between integrative compatible pluralism and non-integrative compatible pluralism.

11.2.2 A Complication: Levels and Pluralism

I propose to refine further Mitchell's analysis by introducing the notion of levels of investigation into the picture (this is suggested by Mitchell herself, in Michell et al. 1997; see also Mitchell and Dietrich 2006). Levels have a long history in philosophy of science (see Oppenheim and Putnam 1958 and Kim 1998). They typically refer to the organization of reality into a hierarchical framework in which the objects at higher levels are constituted by objects at lower levels, in a part/whole relation (Wimsatt 1976; for a slightly different account of levels, see Wimsatt 1994). It is sometimes thought that the different scientific disciplines have different levels of

---

[10] Critics of Mitchell's position have characterize her position, on the basis of her use of examples of the kind mentioned, as a form of "modest pluralism" for her view about the disunity of science "seems to reduce to … monism because it is consistent with the idea that for every particular phenomenon, there is a single, best account" (Kellert et al. 2006, p. xii). More "radical" forms of pluralism are possible in which explanations are not expected to resolve into monism. See for instance, Fehr (2006) on explanations of the evolution of sex and Longino (2006) on the study of human behavior.

[11] Longino is quite clear that this form of compatible pluralism does not lead to integration: "If their partiality is accepted, each approach can be seen to produce some knowledge of behavior by answering the questions distinctive of it with methods that are also distinctive. But none of the approaches can yield a complete account … Each approach can produce partial knowledge. In concert, *they constitute a nonunifiable plurality of partial knowledges*" (2006, p. 127; my emphasis).

reality as their objects, but that is not an accurate rendering of the relations between disciplines and levels. Physics, for instance, studies atoms as well as galaxy and biology cells as well as colony of individuals.

Now, it is often the case that scientific theories at different levels of investigation account for different sets of phenomena. Theories in molecular biology are concerned with the properties of cell components, while physiology focuses on the properties of organs. However, it also happens that theories at different levels of investigation attempt to account for the same set phenomena. In such cases, theories at different levels of explanation might either stand in a relation of competition or be compatible. It is thus useful to distinguish *intra-level pluralism* (when the pluralism in confined to theories at one level of investigation) and *inter-level pluralism* (when the pluralism is between theories at more than one level of investigation). We then have the following typology of pluralism:

- *Interlevel competitive pluralism*: In interlevel competitive pluralism, theories at different levels of investigation are in competition for explaining the same phenomena. For instance, (now discredited) psychoanalytical theories of the etiology of autism once stood in competition with genetic theories. For a long time, the common wisdom has been, at least among reductionist philosophers, that either some lower-level theory would eliminate its competitors at higher levels or that it would reduce them (e.g., Churchland 1986). This position is motivated by the belief that theories at lower levels are more explanatory fundamental in the explanation of higher levels phenomenon (that genes are what explain most aspects of the phenotype of autism, for instance).
- *Interlevel compatible pluralism*: In this form of pluralism, theories at different levels of explanation are thought to explain the same phenomena, but their explanations are taken to be compatible. In a common kind of interlevel compatible pluralism, we have a happy division of labor between theories: Theories at different level focus on different causes or aspects of a phenomenon. But in other case, scientists think of their approaches as in competition, though closer considerations of the approaches show that the theories are not in competition, even if they cannot be unified. As we mentioned earlier, Longino (2006) notes that different approaches, like behavior genetics, neurobiology or socio-environmental approaches, studying human behavior all ask different questions using different methods to answer them. According to her, the multiplicity of approaches is here to stay as they offer only a partial glimpse of the factors determining behavior. As we will see later (Sects. 11.3 and 11.4), these are not the only kinds of interlevel compatible pluralism.
- *Intralevel competitive pluralism*: In this form of pluralism, theories at the same level of investigation are in competition for the explanation of the same phenomena. The issue of this competition should be the elimination of all, but one theory (the true one, for instance).
- *Intralevel compatible pluralism*: In this form of pluralism, different theories at the same level are seen as providing complementary explanations of the same phenomena.

We can use the proposed typology to understand the alternatives to Nagelian reductions. Bickle's proposal is a form of interlevel competitive pluralism where pluralism is temporary (it should give way to explanatory monism) and where the explanation of the phenomena studied will come, not from the level just below the phenomenon, but from far below (at the molecular level). My position concerning certain disciplines is, a form of interlevel compatible pluralism. But as I mentioned, compatible pluralism has many guises. In the following section, I will describe a form of interlevel compatible pluralism that has not received enough attention. I argue that the form of interlevel pluralism instantiated by the cognitive neurosciences of racial prejudice is of a *sui generis* kind.

11.2.3 Integration and Pluralism

Interlevel pluralism is sometimes compatible with some kind of unification. I use the term "integration" to contrast this form of unification with the kind of unification conceived by proponents of the Unity-of-Science project or by contemporary reductionists such as Bickle. By contrast to the latter kind of unification, the integration of two theories entails neither the elimination of one of them nor the identification of the explanatory resources of one to the explanatory resources of the other. Rather, the integration of two theories consists in understanding the specific contribution of each theory to the explanation of some set of phenomena. For instance, integrating two theories might consist in showing that different theories focus on different causes of a phenomenon and how these different causes interact with each other.

The fact that theories can be so integrated is not always obvious. It sometimes takes explicit argument to convince scientists to consider another discipline as relevant to the explanation of the phenomena they are interested in. This was the case with the inclusion by Tinbergen of development in the central questions of ethology (on this point, see Griffiths 2004). I encountered a similar situation in some quarters of psychology. In a previous paper (Machery and Faucher 2005a), I focused my attention on two approaches of racialism (the disposition to classify people into races) that looked for many to be in conflict over explanatory priority (if not explanatory adequacy): the social constructionist approach and the cognitive *cum* evolutionary approach.[12] Most social constructionists believe that the concept of race is a pseudo-biological concept that results from some specific historical circumstances and that it has been used to justify the unequal treatment of specific groups of people. Proponents of the cognitive *cum* evolutionary approach, though they do not believe that races exist either, believe that racialism results from

---

[12] To be more precise, in this case, one should note that the conflict was between theories instantiating different approaches. I will talk about a conflict of approaches or traditions of research to make short.

the working of some specific cognitive system that had a particular evolutionary history. The tenets of these two research traditions has often been judged inconsistent (see our 2005a, for such claims): The social constructionists are vague about the psychological mechanisms responsible for racialism, but their position seems to imply a domain-general form of social learning, whereas the cognitivists *cum* evolutionists argue that racialism is a by-product of dedicated mechanisms. As a result, there has been little contact between the proponents of these two approaches. This is unfortunate, for both approaches have plausibly some empirical evidence and some theoretical insights to contribute to a full understanding of racialism.

I proposed a pluralist solution to this state of affair. One element of my solution (but not the only one, see Machery and Faucher 2005b) was that it was necessary to provide a framework in which data and insights from both field could be integrated. The reason why I thought integration was necessary was that both theories were not really in conflict (so that we were not facing a case of competitive pluralism), but that they were interested in different aspects of the same phenomena: social scientists were interested in the social causes of diversity while evolutionary psychologists were interested in the underlying stability of racial representations (the essentialism underlying racial categorization, for instance). The problem was that theorists from both camps were tempted to give very little weight to the factors that were important for the others (explicitly or implicitly, i.e. by paying only lip service to the other factors). In the best case scenario, theorists of both camps were recognizing that they were in a situation where integrative pluralism should have been actively pursued, but were not doing anything to integrate the different approaches, treating the other one as a competitor. But once it can be shown that the diversity that interests the social scientists can be accommodated by cognitive mechanisms that are not domain-general and once it is shown that by themselves these cognitive mechanisms cannot explain diversity, but that one has to make reference to social and cultural factors to do so, the door is open to an integrated theory of racialism (which I provided, see Machery and Faucher 2005b).

Integrative pluralism is not always that hard to produce: sometimes, scientists are seeing clearly the advantages of integration and are actively working for its accomplishment. The case I would like to consider is the relation between social psychology and neurosciences. In that domain, researchers are seeing the need and explicitly calling for the integration of both disciplines.[13]

---

[13] For instance, Matthew Lieberman wrote that "although the social sciences and neurosciences have been hugely successful enterprises in their own right, *there is a sense that we can now build an intellectual superhighway between them* that will allow us to catalyze the insights from both into a new kind of science that will yield important insights into the basic nature of the human mind . . . " (2006, p. 1; my emphasis).

11.3 Racialism in Social Psychology and Neurosciences

In recent years, a new subdiscipline called "social cognitive neurosciences" (see for instance, Ochsner and Lieberman 2001) has emerged, combining results, methods and tools of many disciplines (such as social, cognitive and developmental psychology, neuropsychology, and sometimes even computer science). In terms of goals,

> [s]ocial cognitive neuroscience has many of the same goals as social psychology in general, but it brings a different set of tools to bear on those scientific goals. (Lieberman, ms)

Indeed, social cognitive neurosciences is basically social psychology using the toolbox of neurosciences, such as fMRI and event-related potentials among others.[14]

Since the turn of the twenty-first century, equipped with these new tools, social psychologists have produced several functional imaging studies investigating the neural basis of psychological capacities implicated in social cognition. For instance, researchers have identified the neural correlates of empathy (Decety 2007), of morality (Green et al. 2001), of understanding of other minds (Frith and Frith 2007) or of cooperation and punishment (Seymour et al. 2007). Work on race and prejudice have not escaped this trend. In this section, I would like to present some of the canonical examples of work done in cognitive neurosciences of racial prejudice. In the next section, I will use these examples (and others) to describe the nature of the relationship between social psychology and neurosciences in this new discipline.

11.3.1 Social Psychology of Prejudice

For now more than 50 years (since Allport's original work, 1954), social psychologists have been trying to explain the origin of many forms of prejudice, including racial prejudice. In that period, social psychology had time to go through important changes concerning its main constructs (Fiske 2000). I am not interested in describing those changes today, but in the current state of the discipline. Before going further, I would like to say a few words concerning how social psychologists are conceiving prejudice.

---

[14] Cacioppo and Bernston defined social psychology as "the scientific study of social behavior, with an emphasis on understanding the individual in a social context. Accordingly, social psychologists study a diverse range of topics ranging from intrapersonal processes shaped by or in response to others, such as the self, attitudes, emotions, social identity, normative beliefs, social perception, social cognition, and interpersonal attraction; to interpersonal processes such as persuasion and social influence, verbal and nonverbal communication, interpersonal relationships, altruism, and aggression, to group processes such as social facilitation, cooperation and competition; . . . " (2006, p. 91). In the late 1970's, social cognitive psychology emerged as a subfield of social psychology, focusing on information-processing accounts of the phenomena to be explained. With the growing success of cognitive neuroscience (Gazzaniga 1995), it was a matter of time before a discipline (or a sub-discipline) like "social cognitive neuroscience" would emerge.

Under the heading “prejudice,” social psychologists have included three phenomena that are often, but not always correlated: stereotype, prejudice per se and discrimination. Studies of prejudices typically look at one of these three phenomena: the cognitive aspect (*stereotype*, for instance, “all members of group X are lazy”), the emotional or attitudinal aspect (*prejudice*, for instance, “I hate people of group X because they are lazy”) and the behavioral aspect (*discrimination*, for instance, “I am not hiring a member of X, because I think or feel that all members of X are lazy”). Stereotypes are typically responsible for the biases in information gathering or in memory that are typical of racial cognition (for a good summary of these effects, see Jones 2002), while prejudice is considered to be responsible for the affective quality of the reaction to outgroup or ingroup members. Both stereotypes and prejudices are thought to have effects on behavior, though as we will see later, their effects might be different.

One question that has become of crucial interest for social psychologists in recent years is the question of “implicit prejudice”.[15] One reason social psychologist have become interested in implicit prejudice is that many people nowadays consider the expression of racism to be politically incorrect. Therefore, investigation of explicit prejudice is open to lie and self-deception on the part of subjects. It is also possible that people are unconsciously influenced by the prevalent culture and sincerely express their beliefs when they claim to be non-prejudiced, while being in fact prejudiced when interacting with people of other races. The egalitarian ideals would then be a superficial veneer covering prejudices that could insidiously have negative effects in interracial relationships. So,

> [t]o measure pure automatic racial bias, unaltered by participants’ social desirability concerns, some researchers place participants in situations where they have less control over their responses or less knowledge about what their responses imply. (Eberhardt 2005, p. 182)

To achieve this end, social psychologists have used tools like lexical priming that allows them to measure *implicit prejudice*. The rational behind this technique is that by

> Presenting a representation of a stereotyped group (e.g. a category relevant word such as “black”) . . . stereotype-relevant knowledge [will be activated] so that immediately subsequent processing of stereotype-relevant information [will be] faster than processing of non-stereotype-relevant information (e.g., Dovidio et al. 1986). (Wheeler and Fiske 2005, p. 60; for a description of the technique used by psychologists, see Kelly and Roedder 2008)

*

Social psychologists have come to an agreement on a definition of the central constructs of their field (stereotypes, prejudice, discrimination). They agree to a large extent on the way to measure some of them (implicit prejudice through lexical priming, for instance). They have also come to postulate that prejudice against

---

[15] As Cunningham et al. put it: “[I]mplicit prejudice can . . . be defined as the automatic cognitive association between a social group and negative evaluation” (2004, p. 1334).

outgroups (including racial groups) is a universal feature of human psychology. Their disagreement concerns mostly how the properties of racial cognition should be explained. Indeed, theories like minimal group theory (or social identity theory, Tajfel and Turner 1986), social dominance theory (Pratto et al. 2006) and others, all predict that social groups will be formed and will be discriminated against but disagree as to which mechanism would be best at explaining racial prejudice. On this issue, there is no unity, but *a form of intralevel competitive pluralism*.

As the following makes clear, social psychology of prejudice fits Darden and Maull's definition of a field, where a field is an entity smaller than a discipline, but bigger than theories that it may contain. According to them, a field consists of "a central problem, a domain consisting of items taken to be facts related to that problem, general explanatory factors and goals providing expectations as to how the problem is to be solved, techniques and methods, and sometimes, but not always, concepts, laws and theories which are related to the problem and which attempt to realize the explanatory goals. A special vocabulary is often associated with the characteristic elements of the field" (1977, p. 44). Cognitive neurosciences of racial prejudice, by trying to build bridges between social psychology of prejudice (a field) and cognitive neurosciences (another field or subdiscipline of neurosciences), could be characterized as an "interfield" enterprise (I am reluctant to call it an "interfield theory" as Darden and Maull are sometimes calling such integration, since the end result of the integration is not a "theory", but more of a bridge—an "intellectual superhighway"—between fields allowing to integrate them further[16]). I will come back to this issue in Sect. 11.4. What should be clear at this point is that the entities that get unified in a process of integration are not necessarily "theories", but could be bodies of knowledge of a different kind (like fields or disciplines). Another thing is that compatible interlevel pluralism is "compatible" with intralevel competitive pluralism (different psychological theories about how the properties of racial cognition should be explained), so both kinds of pluralism are not exclusive.

11.3.2 Cognitive Neurosciences of Racial Prejudice

In order to really understand the type of relationship between fields found in the cognitive neurosciences of racial prejudice, I will present three examples of the kind of work social cognitive neuroscientists have been doing on each aspect of prejudice (stereotype, prejudice and discrimination). I will discuss in the next section how we should understand this work (and some others that revolve around the phenomena presented in this section) in the context of integrative pluralism.

---

[16] As Darden and Maull put it: "… an interfield theory is likely to be generated when background knowledge indicates that relations already exist between the fields, when the fields share an interest in explaining different aspects of the same phenomenon, and when questions arise about that phenomenon within a field which cannot be answered with the techniques and concepts of that field" (1978, p. 50).

1. Social psychologists have postulated that categorization helps us to react faster, therefore more adaptively, to stimuli. As Allport put it: "The human mind must think with the aid of categories … we cannot possibly avoid this process. Orderly living depends on it" (quoted by Ito et al. 2006, p. 189). When it comes to people, we would just use the same domain-general machinery that we use for inanimate objects or non-human animals. One consequence of the process of categorization as applied to human groups is what social psychologists have called the "same race advantage", i.e., the fact that people belonging to an outgroup look more similar to each other (and therefore are harder to discriminate) than people belonging to one's ingroup.[17]

   Neuroscientists extended their inquiry of this phenomenon at the brain level by using what was known of face expertise. Previous work on facial recognition (Kanwisher et al. 1997, Kanwisher 2000; Farah et al. 2000) has suggested that face processing is domain specific and is accomplished by a cerebral structure dedicated to this task (the gyrus fusiform also known as the 'fusiform face area' or FFA). Golby et al. (2001) showed subjects ten black and ten white faces, along with pictures of objects (antique radios). The subjects were asked to remember these stimuli, and their memory was later tested. This memory test yielded a same-race advantage, consistent with previous studies and the

   > imaging data from this study showed that there was greater activation of the FFA when subjects viewed same-race versus other-race faces. This difference was apparent in both the Black and White subjects, primarily in the right fusiform gyrus. Previous studies have found the FFA to be primarily localized in the right hemisphere (e.g., Kanwisher 2000). Golby et al. (2001) also conducted a correlation across subjects in which the compared the magnitude of the FFA response to same-race versus other-race faces with the relative memory advantage for same-race faces. They have found that superior memory for same-race versus other-race faces was significantly correlated with greater activation of the left fusiform gyrus as well as the right hippocampal and parahippocampal gyri, regions known to be important for memory in general (Cohen and Eichenbaum 1993). They proposed that face processing is asymmetric, with the left hemisphere mediating categorical visual processes (i.e., Black vs. White) and the right hemisphere mediating processes involved in individuating faces within a category. (Phelps and Thomas 2003, p. 749).

   One limitation of fMRI technology is that it does not have a very good temporal resolution. To know how quickly race categorization occurs (and this is an important question to help determine how automatic categorization is or how much control one can reasonably have on it), other techniques had to be employed, one of them being event-related potential (ERP). What the use of this method has showed is that categorization is a very early process occurring in the first 100 ms of the perception (Ito et al. 2006).

---

[17] As Susan Fiske puts it: "People accentuate differences between categories and minimize differences within categories (Capozza and Nanni 1986; Tajfel 1970; Taylor 1981). People tag other people by race, gender, and age, so they confuse people within groups and differentiate them between groups (Arcuri 1982; Taylor et al. 1978)." (2000, p. 304)

2. We said earlier that social psychologists have been interested to measure implicit prejudice. Previous studies have identified the amygdala as one of the main structures in charge of early and automatic evaluation of stimuli (Ledoux 1996). Phelps and her colleagues (2000) have

   > . . . used fMRI to examine the relationship between activation of the amygdala and behavioral measures of race bias. During brain imaging, white American subjects were shown pictures of unfamiliar black and white male faces. They were simply asked to indicate when a face was repeated. After imaging, subjects were given a standard explicit assessment of race attitudes (the Modern Racism Scale or MRS; McConahay 1986) and two indirect assessments of race bias: Those subjects who showed greater negative bias on the indirect measures of race evaluation (IAT and startle) also showed greater amygdala activation to the black faces than to the white faces. This correlation between the indirect measures of bias and amygdala activation was not observed when the amygdala response was compared to the MRS, the explicit test of race attitudes. (Phelps and Thomas 2003, p. 752)

   There are some debates on how to interpret the amygdala activity. Phelps and her colleagues think that its activation is due to lack of familiarity and not necessarily to fear of people belonging to an outgroup. They have conducted a study where they showed faces of familiar Black and White individuals. They found that white American subjects are not showing any stronger amygdala activity in response to the presentation of faces of familiar members of both groups.
3. Since we might express prejudice without being aware of it, one question of interest to social psychologists bears on the possibility to attenuate early automatic prejudice (and thus, reduce subtle forms of prejudiced behaviors). Can we control our prejudice? Are there indirect ways by which we can make automatic prejudice disappear? The question has been addressed by looking at modulation of the neural indicator of prejudice following instructions given to subjects. For instance, Todorov and colleagues report that

   > . . . The increased amygdala response to African-American faces most reliably occurred when the white participants were engaged in a one-back recognition task, a categorial same/different task, or a gender categorization task (Hart et al. 2000; Phelps et al. 2000), or in a categorial judgement about the faces, whether the person fell into one or another age social category (Harris and Fiske, unpublished; Wheeler and Fiske 2005). When participants were asked to change their social goal and make individuating judgments (What are the person's likes and dislikes) or perform a visual search task (Is there a dot on the face), the increased activation in the amygdala to black faces diminished below significance. Converging evidence from ERP studies also demonstrates that, once the social goals are changed, the processing of the same faces changes within the first 200 ms (Ito and Urland 2003; Ito et al. 2004). (Todorov et al. 2006, p. 78)

   This thus shows that "[s]imple visual inspection, as in the visual search task was not enough to trigger a differential response to an out-group member either in amygdala activity or in stereotype-knowledge activation (Gilbert and Hixon 1991; Macrae et al. 1997)" (Wheeler and Fiske 2005, p. 61). Change in the context in which subjects view a person can affect out-group perception.

11.4 Integration Without Reduction

Now that I have considered some representative tokens of the kind of research done in cognitive neurosciences of racial prejudice, it is time to describe a bit more precisely what are the modality of the integration. As I will argue (Sect. 11.4.3), the integration cannot be characterized as reductive, but is of a much more complicated nature.

11.4.1 Heuristic Model

Neuroscientists' most obvious contribution to the subfield of the cognitive neurosciences of racial prejudice is a neural model of the mechanisms that constitute racial cognition. Construction of this model requires the integration of the knowledge gathered by social psychologists with the knowledge gathered by neuroscientists. Social psychologists' contribution is twofold. Most important, social psychologists provide many of the phenomena that need to be explained (for instance, the same-race advantage or implicit prejudice). Second, they provide various explanatory theories of the phenomena characteristic of racial psychology. For instance, as we saw above, some social psychologists contend that racial classification results from the domain-general mechanisms of classification, a theory rejected by other psychologists working on racialism.

I will come back to the second of these contributions in a moment. Let's consider the first contribution first. Neuroscientists working on racial psychology provide neural models of the constructs (that includes localization and organization of the different brain areas underlying psychological function) mentioned in social psychologists' explanatory theories. For instance, as we saw in the previous section, they provide neural models of the mechanisms involved in face recognition. In order to do this, neuroscientists mainly draw on the resources of other areas of social cognitive neurosciences, such as face perception or the emotional evaluation of stimuli. They also occasionally draw on neuroscientific findings in domains outside social phenomena, such as the research on fear conditioning (Olsson et al. 2005). Social psychologists' and neuroscientists' knowledge is put together to build a 'heuristic model' of the brain parts causally involved in racial cognition. I call this model 'heuristic' because it does not constitute the ultimate goal of social neurosciences, but rather it is a starting point for the neuroscientific inquiry into racial psychology.[18]

---

[18] The same process of construction of a "heuristic model" is standard in other domain of social cognitive neuroscience (and in cognitive neuroscience). As Ochsner and Lieberman (2001) put it: "When relatively little is known about the neural systems involved in a given form of behavior or cognition, initial studies serve more to identify brain correlates for those phenomena than to test theories about how and why the phenomena occur. This has been the case for many areas

Why construct a neural model of racial cognition and prejudice? Why spend so much time trying to figure out which parts of the brain light up when it executes a particular task? The very same people who use imagery sometimes downplay its role in understanding racial cognition. For instance, Phelps and Thomas wrote:

> This study advances our understanding of the role of FFA and extends it to the processing of race group information. *However, these results do not teach us anything about the behavior of recognizing same-race versus other-race faces that we did not know previously from behavioral work*. We already knew that this advantage exists, is stronger in white than in black Americans and varies across individuals. Knowing how this behavior is represented in the brain does not change these facts. (Phelps and Thomas 2003, p. 750; my emphasis)
>
> .... *it is unclear what we have learned about the behavior of race bias that we did not know before identifying these neural mechanisms*. It may be appealing to extrapolate from these neuroimaging results that such measurements may be able to detect biases that individuals are unwilling to admit, but as this study indicates, behavioral tests are already able to detect such biases. (Phelps and Thomas 2003, p. 752; my emphasis)

How should we understand those statements? Are the results from imagery merely redundant? What's the point of going down to the level of the brain if in the process nothing new is understood or explained? Is imagery just some kind of expansive lite-brite for retarded grown-up, as a researcher once put it?

I think that we have to understand the previous statements as a way to say that the work in neurosciences depends on previous work and theory from social psychologists.[19] Consistent with the description of the heuristic model above, neuroscientists seem to recognize that social psychology has some kind of priority over neurosciences. But is this the kind of historical priority, without any explanatory priority, that characterizes Bickle's view of the relation between higher-level theory and lower-level theory. Is psychology or cognitive science just proving a heuristic to explore the depth of the brain? Is there a hint that they will eventually disappear when cellular/molecular science will succeed? From what we saw in the previous section, this would be an incorrect reading of what is going on in cognitive neurosciences of racial prejudice. In what follow, I will discuss the relationships between theories and fields in cognitive neurosciences of racial prejudice. As I will argue, the best way to conceptualize these relationships is not through reduction but through integration.

---

of cognitive neuroscience research as well. *Ultimately, it will be important to move beyond brain-behavior correlations, but that can only happen when researchers in the fields have built a baseline of knowledge about the brain systems underlying specific types of social or emotional processing* ... Once one has an idea of which processes a brain area carries out, *one can make use of that knowledge to test hypotheses about the involvement of those processes in a given behavior*" (p. 725 and 729; my emphasis).

[19] Neuroscientists seems to acknowledged the priority of social psychology. For instance, Todorov and his colleagues write: "Brain imaging and event-related potential (ERP) studies on race perception have relied on the rich literature in social psychology as a watershed for exploring the neural dynamics involved in stereotyping, prejudice, and other forms of outgroup perception" (Todorov et al. 2006, p. 78).

11.4.2 The Many Roles of Neurosciences in the Cognitive Neurosciences of Racial Prejudice

The contribution of scientists working in the social cognitive neurosciences of racial prejudice is manifold.

(1) One obvious contribution that neuroscientists have done to social psychology, as mentioned earlier, is *localizing psychological constructs in the brain*. Equipped with that knowledge, one can use the activation of certain parts of the brain as indicators of the presence (or psychological reality) of psychological phenomena:

> As more and more is learned about the precise functions of different regions of the brain it may be possible to infer some of the mental processes that an individual is engaged in just from looking at the activity of their brain. (Lieberman, ms)

An example of such a role of the activation of brain regions is the use of the activation of the amygdala as a indicator of implicit negative evaluation I talked about in the second example in Sect. 11.3.2.

(2) The integration of social psychology and neurosciences in the domain of racial psychology allows also for *the discovery of new phenomena*. These new phenomena can either be *psychological*—i.e., as a first approximation, they involve information-processing notions—or strictly *neural*—for instance, they might consist of the localization of some mechanisms. New phenomena can be discovered because the heuristic model allows for *the use of neuropsychological experimental techniques* (primarily brain imagery, but also evoked potential) to extend the range of questions that can be asked concerning the psychological phenomena.

For instance, once the heuristic model is in place, one can study whether the race of a face influences the early stages of face perception or only the later stages of perception or whether it is possible to control the effects of automatic prejudice. For, brain imagery can be used to study the activation of the brain systems involved in perception or of the brain systems involved in self-monitoring. As we saw in Sect. 11.3.2, race seems to influence the perceptual processing of faces and, in certain case as revealed by the use of ERP, this influence manifests itself very early on. In this case and others, the combination of the heuristic model and the investigative tools offered by neurosciences makes it possible to answer questions which might have not been possible with the standard psychological experimental means.

Knowing that a brain structure is implicated in both racial and non-racial cognition makes it also possible to formulate hypotheses about unnoticed psychological phenomena.[20] This is, for instance, what Olsson and colleagues did:

---

[20] Thus, we disagree with Phelps and Thomas above: The integration of neuroscience and social psychology *has* led to the discovery of new psychological phenomena.

> Recent studies have observed that race bias and fear conditioning may indeed rely on overlapping neural systems, suggesting a potential link in mechanism and the opportunity to use classical fear conditioning as a model for aversive learning in a socio-cultural context. (Olsson et al. 2005, p. 785).

The fact that the same structure was involved in fear conditioning led them to look whether some well-known non-social psychological phenomena, like "prepared learning", exist also in a social context. They wondered if the facility to learn a fear response and the difficulty to extinguish the learned response that is thought to be characteristic of fear-relevant stimuli (e.g., snakes and spider)[21] could also be found for members of outgroups. And indeed, they found that it was the case. Subjects learn to fear members from other races faster than members of their own race and this fear is harder to extinguish than for member of their own race. Note that the phenomenon found is psychological: Although the experiment was inspired by neuroscientists' finding that the same neural structure is involved in fear conditioning and in race bias, the methods used by Olsson and colleagues were traditional psychological methods.

(3) Neurosciences have also *lent support for specific psychological theories* in social psychology. For instance, Phelps and Thomas write:

> Studies examining how the brain processes race information have *provided support* for psychological theories concerning the same-race advantage for recognition and the dissociation between direct and indirect measures of race group evaluation. As these studies indicate, a good understanding of the potential contributions of brain imaging can help us discover the structure and organization of a behavior. (Phelps and Thomas 2003, p. 754; my emphasis)

and

> But theory-based accounts of psychologically meaningful brain regions allied to responses of social importance could *provide encouraging evidence for existing theories* (i.e. dual process theories, as just noted) and could facilitate theory development. For example, cross-racial identification apparently links with emotion centers of the brain, which fits together with early indications that prejudice may predict discrimination better than stereotypes do. (Fiske 2000, p. 314; my emphasis)

Neurosciences can support one psychological theory over another, by showing that the neural correlates of the psychological constructs mentioned by the first one, but not the neural correlates of the psychological constructs mentioned by the second one, are causally involved in the production of the relevant psychological phenomena (for a similar discussion, see Willingham and Dunn 2003). *Pace* Bickle, because neurosciences vindicate specific psychological theories, it not the case that neurosciences render psychological explanations otiose and that the latter should be eliminated in favor of (more fundamental) lower-level explanations.

(4) I mentioned earlier that data from other domains of social cognition have contributed to the elaboration of a model of the prejudiced brain (e.g., from

---

[21] See Öhman et Mineka (2003) for a discussion of "prepared learning" for fear-relevant stimuli.

the study of face recognition), but work on the prejudiced brain can also contribute to a better understanding of non-racial phenomena like face perception.

For instance, Gauthier et al. (2000) have proposed that the FFA is not a structure specialized in face recognition. They propose that it is rather specialized in something more general, such as the recognition of objects at the individual level. Gauthier et al. made their point by showing that experts in different fields (like in ornithology) had their FFA activated when they looked at different individuals within their field of expertise.

Using socially charged stimuli, like Black and White faces, Golarai and colleagues

> … exposed White participants both to faces whose features were intact and to faces whose features were moved to random locations on the face (i.e., scrambled). Golarai and colleagues found no difference in fusiform face activation triggered by same-race intact and scrambled face. In contrast, fusiform face activation was significantly reduced for Black intact faces—not only when compared with White intact faces but, amazingly, when compared with White scrambled faces as well. Because it is implausible that participants had more previous exposure to scrambled White faces than to intact Black faces, these results suggest that the fusiform areas may be (a) more sensitive to racial category membership than to the familiarity of the face and (b) more sensitive to racial category membership than to fundamental changes in the structural integrity of the face. (Eberhardt 2005, p. 185).

So using a well-established paradigm developed in the cognitive neurosciences of face-recognition and applying it to socially charged stimuli, Golarai and colleagues have introduced a further complication that face-processing models will have to account for. In this case, the cognitive neurosciences of racial prejudice is producing facts that (psychological) theories from other sub-fields will have to explain. So not only psychological theories not otiose, but in conjunction with neurosciences, they provide new data to explain for psychological theories in other related domains (furthering the need for psychological explanation).

(5) Exchanges between levels can be even more complicated. A closer look at previous theories of social psychology and neurosciences can inspire psychologists and neuropsychologists to distinguish previously confused phenomena. For instance, Amodio and Devine (2006) have argued that the neurosciences of racial cognition have failed to distinguish the cognitive and the affective components of prejudice. To verify the validity of these constructs, they developed some behavioral tests aimed at showing that stereotyped cognition and racial prejudice have different effects. The purpose of such work is to provide a better conceptual background to conduct neuro-imagery. In this case, the source of change is not neurosciences—the psychological theory is not modified in order to be more coherent with the lower-level neuroscientific theory, a case of bottom-up corrective pressures that new wave reductionists think is typical of change in a reductive relationship—but the source of change is internal to psychology.

*

So, what is the nature of the priority of social psychology over neurosciences? First, there is obviously *a historical priority*. Before the development of the cognitive neurosciences of racial cognition, social psychologists have identified phenomena and developed explanatory theories that are relevant for neurosciences. More important, for three reasons, it is *an epistemological priority*. First, social cognitive psychology obviously provides many of the phenomena to be explained. Second, cognitive neurosciences of racial prejudice rely on the heuristic model. The utility of this model itself depends on the validity of the psychological constructs used in social cognitive psychology. If these constructs are not valid, their localization in the brain will be mere artifacts.[22] As the example of Amadio et al. makes clear, this validity is not always established by looking down to the neural substrates, but rather by performing new psychological experiments. Third, and most important, cognitive social psychology provides a level of explanation that is not likely to be eliminated by neurosciences. Indeed, a very important factor in the explanation of why this or that region of the brain will get activated in a certain task is the representations (including stereotypes, norms and associations) that a subject has of the other group or of his own group. Many studies have shown that changing these representations or the way an individual is categorized has an impact on racial phenomena, like feeling of fear of disgust (as indicated by differential activation of the amygdala and the anterior insula; Wheeler and Fiske 2005; Harris and Fiske 2006) or "same-race advantage", indicated by different performances on certain psychological tasks (Cosmides et al. 2003). These representations are usually acquired in social and cultural contexts (see Olsson and Phelps 2007). So in a way, the "social" and "cognitive" dimension of "social cognitive neurosciences" is crucial in the explanation of the phenomena and would typically be lost in a reduction to brain systems or to the molecular level. As I said earlier, explanations of many phenomena will typically need to make reference to causal interactions of the mechanism (the brain systems) with the environment, in this case, the social and cultural environment. But these interactions are not brute causal interactions (like, for instance, when the presence of stressors in the environment increases the release of some hormones without cognitive mediation), they are mediated by cognitive representations that cannot be eliminated from the explanation of the phenomena. So, since both cognitive social psychology and neurosciences are influencing each other and constraining each other progress and since the higher cognitive level seems not to be eliminable, a more accurate depiction of what is going on in cognitive neurosciences of racial prejudice is non-reductive pluralist co-evolution of fields aiming at the greater integration (McCauley 1996).

---

[22] The model is also as good as the neurological theories it is build on. For instance, it might be tempting to interpret amygdala activation in seeing other faces as suggesting that fear is a component of stereotyping. But one might want to resist this temptation by considering that amygdala activation has been associated not only with fear or negative evaluation, but also with positive emotions such as amusement and perception and happiness (Willingham and Dunn 2003, p. 666).

11.4.3 Practical Integration

As we saw in the previous section and in this one, social cognitive psychologists deemed it important to establish connections between the phenomena they are studying and phenomena studied by neuroscientists. Given the state of knowledge and technology, they thought it was possible to build a "heuristic neural model" of racial prejudice, thereby establishing connections between some of the constructs of social psychology and certain brain structures. In none of the cases that I have considered (and I think I have been pretty exhaustive) was there any suggestion that the heuristic model was to somehow replace the psychological model at any point in time or that it was more explanatorily fundamental. By building a bridge between psychology and neurosciences, scientists aimed at answering some questions that could not be answered by the usual psychological means or not with the same degree of certainty. But as I showed, building these bridges, modified both disciplines, allowing them to ask new questions. The cognitive neurosciences of racial prejudice thus illustrate another form of interlevel compatible integrative pluralism then the one I described in Sect. 11.2. In the case I described then, integration (between the cognitive cum evolutionary approach and the social constructionist approach) was produced because theories at different levels were thought to focus on different causes of the same phenomenon. As I made clear, this is not what motivates scientists to cross-disciplinary boundaries in the present case.

I do not want to argue that all integration should be alike. It is possible after all (though dubious), that the cases described by Bickle are better understood as integration by reduction. But, as we saw, unification by reduction falls short of describing all forms of integration found in science. If one accepts my depiction of what is going on in certain (hot and well funded) corners of social cognitive neurosciences, it becomes obvious that we will need a larger and more inclusive framework to think the issue of integration and unity.

Following Grantham (2004), we proposed to think of unity as "interconnection" where the types of interconnections that can produce integration are many.[23] One can indeed distinguish between fields (like social psychology and social neurosciences) as being *theoretically* integrated or *practically* integrated (143). Fields are theoretically integrated when the integration is produced between some of the theories contained in the respective fields. Fields are practically unified, when the unification is accomplished between elements other than theories, for instance methods. Theoretical integration includes the cases of *theoretical reduction* described by Bickle, but also cases of *conceptual refinement*—which, as we saw are not necessarily bottom-up—or the cases *of extension of explanation*—when a phenomenon at one level cannot be explained at that level, but is explained at a lower level (Kitcher 1984; Gray Hardcastle 1992). Fields are *practically* integrated when

---

[23] Grantham talks of "unity" but following my earlier remark, I will rather speak in term of integration.

the relation between them is of a *heuristic nature*—the theories of one field guide the generation of new hypotheses in the other field—when it bears on *confirmation*—data of one field are used to confirm the hypotheses from another field—or when it is of a *methodological nature*—the methods of one field are used in another field. The case of the cognitive neurosciences of racial prejudice is clearly a case of *practical integration* with some (minimal) non-reductive theoretical integration. Indeed, the integration proposed by this subfield requires minimal changes in the central explanatory theories of both fields.

In light of what I just said, it appears that we must refine the typology proposed in Sect. 11.2. Interlevel compatible pluralism can include cases of theoretical integration and cases of practical integration, as well as a mix of both. Some cases of theoretical integration will be motivated by the fact that the causal factors responsible for a phenomena are studied by different theories; some other cases will be motivated by the fact that explanation of a phenomena cannot be provided by higher-level theories. Some other forms of integration will be mostly practical, as when a discipline would use the methods of another discipline to validate its own method (think of how neurosciences used lesion work in the course of validation of the results of brain imagery; see Bechtel 2004 and Wylie 2000). Finally, some fields will illustrate a mix of both forms of integration (and here it can be to different degree). From that picture we can see that the possible connections between fields cannot be reduce to 'reduction' whichever way you conceive it.

11.5 Conclusion

Most recent discussions of the unity of science have assumed the "unity by reduction" thesis. A version of that thesis sees reduction as a derivation of higher-level laws from lower-level laws. Many philosophers believe that in that form, the project of unification is dead. We are at a cross-road. The choice seems to be between a new style of reduction and disunity. Bickle choose the first option. I am leaning toward a third option (I have not considered the second option, but it goes without saying that if I am right, at least part of science is not condemned to disunity). Adopting the "new wave metascience" stance advocated by Bickle, I claimed that the examination of some actual cases of integration in actual science vindicates a much more liberal position on unification, a position much in the spirit of the one some of the fathers of the Unity of Science project had in mind. From a detailed description of the modalities of pluralism in cognitive neurosciences of prejudice, I have concluded (1) that Mitchell's typology of pluralism is on the right track, but should be supplemented (compatible pluralism can be integrative or non integrative, integration can be theoretical or practical or both, intralevel competitive pluralism can coexist with interlevel integrative pluralism) and (2) that Bickle's form of reductionism does not describe at all what is going one in that domain of science. The relations between fields (and theories inside them) in the cognitive neurosciences of racial prejudice are non-reductive and of a more practical nature.[24] Discussions of the unity of science that get their inspiration from science in the wild, as Bickle claims to do, should be open to the idea that in science "unification" could take different forms. Pluralism also rules here!

Acknowledgments Many thanks to Edouard Machery who read and commented several versions of this paper. He is responsible of many of the good ideas in it. Thank you also to Frédéric Bouchard and Pierre Poirier for comments on the last version of the paper. I tried my best to accommodate all their comments. I am also indebted to John Bickle for his comments on a shorter version of this paper.

References

Allport, G. 1954. *The nature of prejudice*. Reading: Addison-Wesley Publishing Co.

Amodio, D.M., and P.G. Devine. 2006. Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology* 91(4): 652–661.

Bechtel, W., and A. Abrahamsen. 2008. *From reduction back to higher levels*. Proceedings of the 30th Annual Meeting of the Cognitive Science Society (pp. 559–564). Austin, TX: Cognitive Science Society.

Beatty, J. 1994. Ernst Mayr and proximate/ultimate distinction. *Biology and Philosophy* 9: 333–356.

Bechtel, W. 2004. The epistemology of evidence in cognitive neuroscience. In *Philosophy and the life sciences: A reader*, ed. R. Skipper, C. Allen, R.A. Ankeny, C.F. Craver, L. Darden, G. Mikkelson, and R. Richardson. Cambridge: MIT Press.

Bechtel, W., and A. Hamilton. 2007. Reductionism, integration, and the unity of the sciences. In *Philosophy of science: Focal issues*, The handbook of the philosophy of science, vol. 1, ed. T. Kuipers. New York: Elsevier.

Bickle, J. 1998. *Psychoneural reduction: The New wave*. Cambridge: MIT Press.

Bickle, J. 2003. *Philosophy and neuroscience: A ruthlessly reductive account*. Dordrecht: Kluwer Academic Publishers.

Bickle, J. 2006. Reducing mind to molecular pathways: Explicating the reductionism implicit in current mainstream neuroscience. *Synthese* 151(3): 411–434.

Cacioppo, J.T., and G. Bernston. 2006. A bridge linking social psychology and the neurosciences. 920 In *Bridging social psychology: Benefits of transdisciplinary approaches,* ed. P.A.M. Van Lange, 921 91–96. Mahwah: Lawrence Erlbaum Associates. 922

Cat, J. 2007. Scientific Unity. In *Stanford encyclopedia of philosophy*. http://plato.stanford.edu/entries/scientific-unity/.

---

[24] Some might say that this is because the theories in each domains are not ripe for reduction or that the theories of cognitive social psychology are false (and therefore could not be derived or reduce to true theories). But one could argue the other way around and say that neurosciences of memory are reductionist because they are not ripe. Bechtel and Abrahamsen (2008) have advocated such a view showing that reduction, in certain case, is a first step that is then to be complemented by an account that considers the processes that were left behind in the reductionistic quest. My position is that in the case I studied, it is very unlikely that we will be able to do away with the cognitive and social levels of explanation.

Cat, J., H. Chang, and N. Cartwright. 1996. Otto Neurath: Politics and the unity of science. In *The disunity of science: Boundaries, contexts, and power*, ed. P. Galison and D.J. Stump, 347–369. Stanford: Stanford University Press.

Chomsky, N. 1965. *Aspects of the theory of syntax*. Cambridge: MIT Press.

Churchland, P. 1986. *Neurophilosophy: Toward a unified science of the mind-brain*. Cambridge: MIT Press.

Cosmides, L., J. Tooby, and R. Kurzban. 2003. Perceptions of race. *Trends in Cognitive Sciences* 7(4): 173–179.

Craver, C.F., and L. Darden. 2001. Discovering mechanisms in neurobiology: The case of spatial memory. In *Theory and method in neuroscience*, ed. P.K. Machamer, R. Grush, and P. McLaughlin, 112–137. Pittsburgh: University of Pittsburgh Press.

Creath, R. 1996. The unity of science: Carnap, Neurath and beyond. In *The disunity of science: Boundaries, contexts, and power*, ed. P. Galison and D.J. Stump, 158–169. Stanford: Stanford University Press.

Cunningham, W.A., J.B. Nezlek, and M. Banaji. 2004. Implicit and explicit ethnocentrism: Revisiting ideologies of prejudice. *Personality and Social Psychology Bulletin* 30(10): 1332–1346.

Darden, L., and N. Maull. 1977. Interfield theories. *Philosophy of Science* 44: 43–64.

Decety, J. 2007. A social cognitive neuroscience model of human empathy. In *Social neuroscience: Integrating biological and psychological explanations of social behavior*, ed. E. Harmon-Jones and P. Winkielman, 246–270. New York: Guilford Press.

Dupré, J. 1993. *The disorder of things: Metaphysical foundations of the disunity of science*. Cambridge: Harvard University Press.

Dupré, J. 1996. Metaphysical disorder and scientific disunity. In *The disunity of science: Boundaries, contexts, and power*, ed. P. Galison and D.J. Stump, 101–117. Stanford: Stanford University Press.

Eberhardt, J.L. 2005. Imaging race. *American Psychologist* 60(2): 181–190.

Farah, M.J., C. Rabinowitz, G.E. Quinn, and G.T. Liu. 2000. Early commitment of neural substrates for face recognition. *Cognitive Neuropsychology* 17: 117–124.

Faucher, L. 2006. What's behind a smile: Commentary on Schaffner. *Synthese* 151(3): 403–409.

Fehr, C. 2006. Explanations of the evolution of sex: A plurality of local mechanisms. In *Scientific pluralism*, Minnesota studies in philosophy of science, vol. XIX, ed. S.H. Kellert, H.E. Longino, and C.K. Waters, 167–189. Minneapolis: University of Minnesota Press.

Fiske, S. 2000. Stereotyping, prejudice, and discrimination at the seam between the centuries: Evolution, culture, mind, and brain. *European Journal of Social Psychology* 30: 299–322.

Fodor, J.A. 1974. Special sciences (or: The disunity of science as a working hypothesis). *Synthese* 28: 97–115.

Fodor, J.A. Look. *Review of consilience: The unity of knowledge*. In The London Review of Books, ed. E. O. Wilson, 1988. 29 Oct (online: http://www.lrb.co.uk/v20/n21/fodo01_.html).

Frith, C., and U. Frith. 2007. Social cognition in humans. *Current Biology* 17(16): R724–R732.

Gauthier, I., P. Skudlarski, J.C. Gore, and A.W. Anderson. 2000. Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience* 3(2): 191–197.

Gazzaniga, M.S. (ed.). 1995. *The cognitive neurosciences*. Cambridge: MIT Press.

Glennan, S. 2002. Rethinking mechanistic explanation. *Philosophy of Science* 69: S342–S353.

Golby, A.J., J.D.E. Gabrieli, J.Y. Chiao, and J.L. Eberhardt. 2001. Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience* 4(8): 845–850.

Grantham, T. 2004. Conceptualizing the (dis)unity of science. *Philosophy of Science, vol.* 71: 133–155.

Gray Hardcastle, V. 1992. Reduction, explanatory extension, and the mind-brain sciences. *Philosophy of Science* 59: 408–428.

Green, J.D., R.B. Sommerville, L.E. Nystrom, J.M. Darley, and J.D. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293: 2105–2108. Sept. 14th.

Griffiths, P.E. 2004. Instinct in the '50s: The British reception of Konrad Lorenz's theory of instinctive behaviour. *Biology and Philosophy* 19(4): 609–631.

Hacking, I. 1996. The disunities of science. In *The disunity of science: Boundaries, contexts, and power*, ed. P. Galison and D. Stump, 38–74. Stanford: Stanford University Press.

Harris, L.T., and S.T. Fiske. 2006. Dehumanizing the lowest of the low: Neuro-imaging responses to extreme outgroups. *Psychological Science* 17: 847–853.

Ito, T.A., G.R. Urland, E. Willadsen-Jensen, and J. Correll. 2006. The social neuroscience of stereotyping and prejudice: Using event-related potentials to study social perception. In *Social neuroscience: People thinking about thinking people*, ed. J.T. Cacioppo, P.S. Visser, and C.L. Pickett, 189–208. Cambridge: MIT Press.

Jones, M. 2002. *Social psychology of prejudice*. Upper Saddle River: Prentice Hall.

Kanwisher, N., J. McDermott, and M. Chun. 1997. The fusiform face area: A module in human extrastriate cortex specialized for the perception of faces. *Journal of Neuroscience* 17: 4302–4311.

Kanwisher, N. 2000. Domain specificity in face perception. *Nature Neuroscience* 3(8): 759–763.

Kellert, S.H., H.E. Longino, and C.K. Waters. 2006. The pluralist stance. In *Scientific pluralism*, Minnesota studies in philosophy of science, vol. XIX, ed. S.H. Kellert, H.E. Longino, and C.K. Waters, vii–xxix. Minneapolis: University of Minnesota Press.

Kelly, D., and E. Roedder. 2008. Racial cognition and ethics of implicit bias. *Philosophical Compass* 3(3): 522–540.

Kim, J. 1998. *Mind in a physical world: An essay on the mind-body problem and mental causation*. Cambridge: MIT Press.

Kitcher, P. 1984. 1953 and all that: A tale of two sciences. *The Philosophical Review* 93: 335–373.

Kitcher, P. 1999. Unification as a regulative ideal. *Perspectives on Science* 7(3): 337–348.

Kosslyn, S.M. 1980. *Image and mind*. Cambridge: Harvard University Press.

Kuhn, T. 1962. *The structure of scientific revolutions*. Chicago: Chicago University Press.

Ledoux, J. 1996. *The emotional brain*. New York: Simon and Schuster.

Lieberman, M. 2006. Social cognitive and affective neuroscience: When opposites attract. *Social Cognitive and Affective Neuroscience* 1(1–2): 1–2.

Lieberman, M.D. 2010. Social cognitive neuroscience. S.T. Fiske, D.T. Gilbert, and G. Lindzey (Eds). *Handbook of Social Psychology* (5th ed.) (pp. 143–193). New York, NY: McGraw-Hill.

Longino, H.E. 2006. Theoretical pluralism and the scientific study of behavior. In *Scientific pluralism*, Minnesota studies in philosophy of science, vol. XIX, ed. S.H. Kellert, H.E. Longino, and C.K. Waters, 102–131. Minneapolis: University of Minnesota Press.

Looren de Jong, H. 2006. Explicating pluralism: Where the mind to molecule gets off the track—reply to Bickle. *Synthese* 151: 435–443.

Machamer, P., L. Darden, and C. Craver. 2000. Thinking about mechanisms. *Philosophy of Science* 67: 1–25.

Machery, E., and L. Faucher. 2005a. Why do we think racially? A critical journey into culture and evolution. In *Handbook of categorizatin in cognitive science*, ed. C. Lefèbvre and H. Cohen, 1009–1033. New York: Elsevier.

Machery, E., and L. Faucher. 2005b. Social construction and the concept of race. *Philosophy of Science* 72: 1208–1219.

McCauley, R. 1996. Explanatory pluralism and the coevolution of theories in science. In *The churchlands and their critics*, ed. R. McCauley. Oxford: Blackwell Publishers.

McCauley, R.N., and W. Bechtel. 2001. Explanatory pluralism and the Heuristic Identity theory. *Theory and Psychology* 11: 736–760.

Mitchell, S.D. 2002. Integrative pluralism. *Biology and Philosophy* 17: 55–70.

Mitchell, S.D. 2003. *Biological complexity and integrative pluralism*. Cambridge: Cambridge University Press.

Mitchell, S.D., and M.R. Dietrich. 2006. Integration without unification: An argument for pluralism in the biological sciences. *The American Naturalist* 168: S73–S79.

Mitchell, S.D., L. Daston, G. Gigerenzer, N. Sesardic, and P.B. Sloep. 1997. The why's and how's of interdisciplinarity. In *Human by nature: Between biology and the social sciences*, ed. P. Weingart, S.D. Mitchell, P. Richerson, and S. Maasen, 103–150. Mahwah: Erlbaum Press.

Nagel, E. 1961. *The structure of science*. New York: Hartcourt, Brace and World.

Ochsner, K.N., and M. Lieberman. 2001. The emergence of social cognitive neuroscience. *American Psychologist* 56: 77–734.

Öhman, A., and S. Mineka. 2003. The malicious serpent: Snakes as a prototypical stimulus for an evolved module of fear. *Current Directions in Psychological Science* 12(1): 5–9.

Olsson, A., and E.A. Phelps. 2007. Social learning of fear. *Nature Neuroscience* 10(9): 1095–1102.

Olsson, A., J. Ebert, M. Banaji, and E.A. Phelps. 2005. The role of social groups in the persistence of learned fear. *Science* 308: 785–797.

Oppenheim, P., H. Putnam, et al. 1958. The unity of science as a working hypothesis. In *Minnesota studies in the philosophy of science*, vol. 2, ed. H. Feigl. Minneapolis: Minnesota University Press.

Phelps, E.A., K.J. O'Connor, W.A. Cunningham, E.S. Funayma, J.C. Gatenby, J.C. Gore, M.R. Banaji. 2000. Performance on indirect measures of race evaluation predicts amygdala activity. *Journal of Cognitive Neuroscience* 12: 1–10.

Phelps, E.A., and L.A. Thomas. 2003. Race, behavior and the brain: The role of neuroimaging in social behaviors. *Political Psychology* 24(4): 747–758.

Pratto, F., J. Sidanius, J., and S. Levin. 2006. Social dominance theory and the dynamics of intergroup relations: Taking stock and looking forward. In W. Stroebe, and M. Hewstone (eds.), *European Review of Social Psychology* 17: 271–320.

Pylyshyn, Z.W. 1981. The imagery debate: Analogue media versus tacit knowledge. *Psychological Review* 88: 16–45.

Queller, D.C., and J.E. Strassmann. 1998. Kin selection and social insects. *Bioscience* 48: 165–175.

Richardson, A.W. 2006. The many unities of science: Politics, semantics, and ontology. In *Scientific pluralism*, Minnesota studies in philosophy of science, vol. XIX, ed. S.H. Kellert, H.E. Longino, and C.K. Waters, 1–25. Minneapolis: University of Minnesota Press.

Sarkar, S. 1992. Models of reduction and categories of reductionism. *Synthese* 91: 167–194.

Seymour, B., T. Singer, and R. Dolan. 2007. The neurobiology of punishment. *Nature Reviews Neuroscience* 8: 300–311.

Tajfel, J., and J.C. Turner. 1986. The social identity theory of inter-group behavior. In *Psychology of intergroup relations*, ed. S. Worchel and L.W. Austin, 2–24. Chicago: Nelson-Hall.

Tinbergen, N. 1963. On aims and methods in ethology. *Zeitschrift für Tierpsychologie* 20: 410–433.

Todorov, A., L.T. Harris, and S. Fiske. 2006. Toward socially inspired social neuroscience. *Brain Research* 1079: 76–85.

Wheeler, M.E., and S. Fiske. 2005. Controlling racial prejudice: Social-cognitive goals affect amygdala and stereotype activation. *Psychological Science* 16(1): 56–63.

Willingham, D.T., and E.W. Dunn. 2003. What neuroimaging and brain localization can do, cannot do, and should not do for social psychology. *Journal of Personality and Social Psychology* 84(4): 662–671.

Wimsatt, W. 1976. Reductionism, levels of organization, and the mind-body problem. In *Consciousness and the brain: A scientific and philosophical inquiry*, ed. I. Savodnik, 202–267. New York: Plenum Press.

Wimsatt, W. 1994. The ontology of complex systems: Levels, perspectives, and causal thickets. *Canadian Journal of Philosophy. Supplemental Volume* 20: 207–274.

Wylie, A. 1999. Rethinking unity as a 'working hypothesis' for philosophy of science: How archeologists exploit the disunities of science. *Perspectives on Science* 7(3): 293–317.

Wylie, A. 2000. Questions of evidence, legitimacy, and the (dis)unity of science. *American Antiquity* 65(2): 227–237.

Chapter 12
Sciences as Open Systems – The Case of Economics*

Vítor Neves[1]

Abstract Unity/disunity and specialization/unification are usually thought of in terms of opposing duals. As such, they have marked the debate about science for a number of years. In this chapter it is argued that connections (and the isolative strategies we adopt) are crucial and that an understanding of sciences as theoretical open systems has the potential to break the barrier between specialization and unification. A suggestion is made to consider open systems of knowledge as the way to manage both our cognitive limits and the ultimate, ontological, interconnectedness of the world. Taking this route, it is shown that the full adoption of an open systems approach in economics will have deep implications for the way the discipline is understood and developed as well as how it relates to the other social sciences.

Keywords Connections • Economics • Integration of social knowledge • Interdisciplinary isolations • Open systems

---

*This chapter builds on (and significantly revises) the paper I presented at the First Lisbon Colloquium for the Philosophy of Science – Unity of Science. Non Traditional Approaches, Lisboa, Teatro da Trindade, 25–28 October 2006.

[1] Vítor Neves is Assistant Professor in the Faculty of Economics of the University of Coimbra, Portugal, where he concluded his doctorate in economics in 2003. He is also researcher at the Centre for Social Studies (CES, University of Coimbra), collaborates with the Centre for Philosophy of Sciences of the University of Lisbon (CFCUL) and is a member of the Stirling Centre for Economic Methodology (SCEME), United Kingdom. His research interests are focused on the history of economic thought, economic methodology, public economics and housing economics. He has been doing research on the meaning of the 'economic'; economics and interdisciplinary exchange; open systems; objectivity in economics; and social costs and social value.

V. Neves (✉)
Faculdade de Economia da Universidade de Coimbra, Av. Dias da Silva,
165, 3004-512 Coimbra, Portugal
e-mail: vneves@fe.uc.pt

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology, and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_12,

12.1 Introduction

> *Utopia is on the horizon.*
> *I get two steps closer, and it moves two steps away. I walk ten steps and it is ten steps further away.*
> *No matter how much I walk, it will never be reached.*
> *So what is utopia for? For this it serves: to walk.*
>
> Eduardo Galeano

Disunity rather than unity seems to characterize better the present situation and the dynamics of contemporary science. The growth of knowledge is still primarily the product of specialized research. New, ever more specialized fields continue to emerge and, in a number of cases, scientific developments lead to fragmentation and disintegration. But it is also true that many important advances have been made as the result of conceptual, theoretical and methodological borrowings and spillovers across the sciences. Several transformations in the direction of unification are detectable within and between specialized sciences. Also, pleas for unity and integration (whatever these mean)[2] continue to be made.

Trends towards both integration and disintegration are observable in the social sciences. This is also the case in economics. Conflicting tendencies and tensions are perceptible in the recent evolution of the economic science, from the economics-centred reductionist movement towards unification known as 'economics imperialism' to the other more pluralist, two-way approaches, which one may accurately describe in terms of import-export trade, mutual inspiration and conceptual/theoretical contamination. Noticeable in this context are interdisciplinary exchanges – both horizontal and vertical[3] – which have moved research frontiers and have given place to innovative, now very dynamic areas of research in economics, such as 'law and economics', socioeconomics, behavioural economics, experimental economics, neuroeconomics, evolutionary economics, econophysics and complexity economics.[4] It is also the case that some of these areas of research may threaten the old disciplinary unity.

---

[2] As Pombo (2006, 29) noted, the idea of "unity of science" means a variety of things, each of these meanings being connected to a different conceptual context and set of problems. A similar view can be found in Hacking (1996), who suggested three different "unities" of science: a *metaphysical belief* (or sentiment) in one scientific world, reality or truth; a *practical precept* to look for the connections between phenomena; and a single standard or *mode of scientific reasoning*.

[3] We call "horizontal exchanges" to interactions taking place between social sciences such as economics, sociology, political science, anthropology, and law. "Vertical exchanges" occur when social sciences take inspiration and seek guidance and insight from lower-level physical, cognitive and life sciences.

[4] For a very interesting discussion of the changes occurring in recent economics, in which it is suggested that one uses an import-export model and looks at the history of economics in terms of a (non-regular) cycle of alternating unity and plurality periods, see Davis (2006).

What does it mean today to speak of the *unity of science* in such a context? Are there prospects – after all the criticisms against Vienna Circle's project on the unity of science – for any kind of successful pursuit of unification within and between sciences?

Science is plural. All we have (and what is reasonable to foresee) are *sciences*. The idea of Science as a Grand Unified Theory (or approach), or the requirement that all sciences be expressed in a common language (e.g. mathematics) or pursued with the same method (*the* scientific method, whatever that means), seem to face insurmountable difficulties. Nevertheless, my belief is that the *utopia of unity* (and the search for coherence, within and across sciences) should be kept at the back of one's mind, even if just as an aspiration, an unreachable ideal which leads us to transgress disciplinary boundaries (in an overall context of 'dialogical encounter' across disciplines – or 'interdisciplinary encounters'[5] – and cross-fertilization) and keeps us looking for *connections* and the *totality*. I claim that one may find a rationale for *unity-in-diversity*[6] and interdisciplinary exchange in an open-systems understanding of social reality and science. My point is that, against the dualist view of science in terms of unity/disunity,[7] of ever more specialized, fragmentary (closed and self-sufficient) sciences opposed to reductionist projects of unification, sciences are to be construed (and developed) as *theoretical open systems*, with *emergent* properties irreducible to those of any one of them or to whatever basic characteristic, language or method/logic of inquiry one may consider.

I maintain that this view of the sciences as open systems has the potential to overcome the above mentioned tension between specialization and unification. The basic assumption is that *systems* and *sciences as systems* should be taken as the units of analysis. In this chapter I take economics as the focal point. My purpose is to show that the full adoption of an open systems approach would have deep implications on how the discipline is understood and developed. Attention will be paid, in particular, to the ever present dilemma between the ultimate interconnectedness of things and the need to decompose our object of study in order to make it manageable and to generate knowledge. The chapter is organized as follows. In the next section the concepts of interdisciplinary isolation and isolative strategies are introduced. In Sect. 12.3 I discuss the problem of missing connections and the challenge this brings to our analyses based on those isolations. In Sect. 12.4 open and closed systems are

---

[5] See http://www.helsinki.fi/filosofia/tint/events/interdisciplinaryencounters.htm.

[6] Or, following Hacking's (1996) terminology, *unity as integrated harmony* of different domains of inquiry rather than *unity as singleness* ("the subsumption of all phenomena under a single principle, law, or language").

[7] Following Dow, dualism may be defined as the propensity to classify concepts, statements and events in terms of all-encompassing, mutually exclusive pairs of either/or categories, with fixed meanings (cf., for instance, 1996; 2002). Yet, as this author noted, although "[m]ethodological discussion has often focused on duals: induction/deduction, rationalism/empiricism, prescription/description, modernism/postmodernism (...) none of these purist positions is actually sustainable" (2002, 167).

distinguished and some proposals to build economics as an open system briefly reviewed. Some conclusive remarks and suggestions for further research will be outlined in the final section.

12.2 Interdisciplinary Isolations and Isolative Strategies

In the process of becoming autonomous disciplines, sciences have to define their respective domains of study, which involves delimiting (and to a certain extent closing) their boundaries. A set of explaining items (the *explanantia*) and a set of explained items (the *explananda*) have to be sorted out. According to Mäki (1992 and 2004), all this entails undertaking *interdisciplinary isolations*.[8] A (theoretical) isolation is defined as a process whereby a set X of entities is assumed to be theoretically removed ('sealed off') from the involvement or influence of another set Y (X, the *isolated field*, and Y, the *excluded field*, exhaust all relevant possibilities). An interdisciplinary isolation may then be thought as one which helps define the domain of a scientific discipline *by delimiting and closing its boundaries*.

Isolations are unavoidable. We have to close our minds to many possibilities (at least temporarily) in order to pay attention to a few (Loasby 2003, 294). Some connections have always to be supposed absent or inoperative.[9] The choice of the isolated field of a science, which includes both its explananda and explanantia, entails exclusions. These are usually based on *omission*, i.e. the excluded entities, features or aspects tend simply to be ignored and fall outside of the scope of the discipline without any mention of them. They become a "field of silence" (Mäki 1992, 335–6).[10]

As what can be explained by a science is constrained by the chosen set of explaining items (Mäki 2004, 322), the interdisciplinary isolations involved in the choice of both the explananda and the explanantia – and the consequent boundary lines drawn – become crucial and controversial matters. Various interdisciplinary isolative strategies are observable in economics as well as in the other social sciences regarding these issues. Important dynamics of dispute may arise both regarding

---

[8] The 'interdisciplinary isolation' conceptual framework, proposed by U. Mäki, will be considered here as a helpful, descriptive device for understanding the process that led to the present situation of compartmentalized and segmented sciences. For a critique of isolation as a method in the social sciences, see Runde (1996) and Lawson (1997).

[9] Isolations, Mäki claims, are ubiquitous: "Every concept, model and theory is based on an isolation of a slice of the things and properties in the world to the exclusion of the rest of what there is" (2004, 321).

[10] In some cases (although these are more uncommon in interdisciplinary isolations) items are explicitly considered but excluded from the isolated disciplinary field by means of some form of *idealization*, usually through the postulation of limits or ideal types (for instance, setting the value of variables to be 0 or $|\infty|$).

the items to explain and the explaining items one science chooses to emphasise. Frequently, these are disputes over different conceptions of the nature of the object of study.

Following Mäki (2004), let us consider a set S of potential *explaining* items – comprising of an isolated subset $S_k{}^- = \{s_1, \ldots, s_k\}$ and a subset of excluded items $S_k{}^+ = \{s_{k+1}, \ldots\}$ – and, similarly, a set M of potential *explained* items also encompassing two subsets – $M_n{}^- = \{m_1, \ldots, m_n\}$, the isolated explananda, and $M_n{}^+ = \{m_{k+1}, \ldots\}$, the excluded potential explained items. It is clear that the choice of the boundary lines between $S_k{}^-$ and $S_k{}^+$, on one hand, and between $M_n{}^-$ and $M_n{}^+$ on the other hand, are of the utmost relevance.

‘Economics imperialists’, for instance, based on a definition of economics as the science of human choice in conditions of scarcity, claim that no area of human action is outside the potential scope of economics. Moreover, the conventional rational choice model, with its origins in economics, is seen as providing “the most promising basis presently available for a unified approach to the analysis of the social world by scholars from different social sciences” (Becker, Nobel Lecture, 1993). The boundary between $M_n{}^-$ and $M_n{}^+$ is thus increasingly pushed forward by means of an employment of the traditional $S_k{}^-$ of mainstream economics to explain traditionally ‘non-economic’ areas such as marriage and family, education, crime or discrimination.

In contradistinction, another Nobel laureate, Ronald Coase, rejects such a view of economics as a toolbox of universal application, saying that “[t]he analysis developed in economics is not likely to be successfully applied in other subjects without major modifications” (1994, 44). He underlines, instead, a conception of economics based on the centrality of the *economy*, conceived of as intermeshed with other parts of the broader social system. In his view, the distinguishing feature of economics is a concern with the study of the “working of the social institutions which bind together the economic system: firms, markets for goods and services, labour markets, capital markets, the banking system, international trade and so on” (ibid, 41), a study which Coase thought “hardly possible” without consideration of the effects of the other parts of the social system on the working of the economic system itself. That is, an enlargement of the *explanantia* of economics, to include items from $S_k{}^+$, is recommended for an explanation of a redefined $M_n{}^-$.[11]

---

[11] As Coase (1994, 46) puts the issue:
the study by economists of the effects of the other social systems on the economic system will, I believe, become a permanent part of the work of economists. It cannot be done effectively by social scientists unfamiliar with the economic system. Such work may be carried out in collaboration with other social scientists but is unlikely to be well done without economists. For this reason, I think we may expect the scope of economics to be permanently enlarged to include studies in other social sciences. But the purpose will be to enable us to understand better the working of the economic system.

12.3 The Problem of the Missing Connections

Important as interdisciplinary isolations might be for the constitution of a science as an autonomous discipline, they imply that some connections become concealed or even inexorably lost. Isolations apply only to our thoughts and representations of the real world; not to the phenomena themselves (Loasby 1999, 14).[12] As with the ocean, we may separate phenomena into parts and label them (the Atlantic, the Pacific, the Indian, etc.). However, the ocean does not stop being a single, immense, interconnected and continuous body of salt water encircling the Earth.[13] Boundaries are a human, artificial construct. The world has no boundaries; it forms a single ontological unity.

It is true that, as already mentioned, we cannot avoid making isolations. As Loasby emphasised, "[k]nowledge grows by division; each of us can increase our knowledge only by accepting limits on what we can know" (ibid, 135). But because it is dispersed and "[b]ecause different people can develop different skills, a knowledge-rich society must be an *ecology of specialists*; knowledge is distributed (. . .) and being distributed it can grow, *provided that it is sufficiently co-ordinated to support increasing interdependencies*" (ibid, 130, emphases added). Once we subscribe to an understanding of social reality as being highly interconnected and organic, reliance on the sufficiency of segmentations that cut "swathes through the recognised interconnectedness of things" (Chick and Dow 2001, 714) may lead us astray. Isolations inevitably exclude more or less relevant connections between the phenomena being studied. Our knowledge, as in the case of the elephant described by the seven blind men, can only be partial and incomplete. Error as such is unavoidable and co-ordination of different contributions to knowledge indispensable. However, and this needs also to be stressed, incompleteness is a source of innovation as well. Incompleteness, as Loasby taught us, "is not only a source of problems; it opens up our systems of thought" (2003, 293):

> Since our representations are always incomplete, innovation is always possible; we can change the set of elements, revise the internal linkages between them or redefine the external connections. Whether contemplating artefacts, processes, structures, sequences, problems or strategies, we are operating in large combinatorial spaces in which there are, in principle, many options for change. (ibid, 301)

Each science (and the knowledge it produces) is a set of connections that grows by making novel connections (which, very often, also involve destroying and substituting other connections) and it forms, with the other sciences, a wider *open system of knowledge*, in the context of which, various connections are established.[14]

---

[12] Of course, there are also *material isolations* occurring in nature, but apart from some very infrequent cases of spontaneous isolations, they are usually the result of human intervention taking place in an experimental context (*experimental isolations*).

[13] On the beautiful Leibnitzian metaphor of science as ocean, see Pombo (2006).

[14] A system, be it closed or open, is a network, a structure with connections – a *set of elements* (things or ideas) linked by a *set of connections* so as to form a coherent whole (cf. Potts 2000).

Since "[l]arge systems typically include sub-systems within which connections are relatively dense, and between which connections are sparse" (Loasby 2005, 59), the challenge becomes *how to decompose the object of our knowledge* so that the co-ordination of the different strands of thought becomes possible in a coherent and consistent manner.

This brings us back to the nature of the interdisciplinary isolations we undertake. There are limits to these isolations. Such limits, as Mäki (2001) suggested, are determined by our conception of "the way the world works", that is, they are subject to an ontological constraint. Rival interdisciplinary isolative strategies (dealing with choices of *what* is isolated and *how*) have an ontological basis. The legitimacy of any isolative strategy is always dependent on our conception of the nature of the reality object of study. Items considered to play a *necessary* or *essential* role in the functioning of the world – "ontic indispensability" – should not be excluded from our systems of thought. Similarly, there are "ontic impossibilities" that should make us cautious regarding isolations that lead to theories which "depict the world in such a way that we have reason to believe that the world *does not* function that way, or, more strongly, that it *cannot* function that way, or, still more strongly, that it cannot function at all, given what we know about it". (Mäki 2001, 385, original emphasis)

12.4 Economics as an Open System

One of the most controversial issues which has long separated several heterodox traditions in economics (as well as other social scholars) from the mainstream of the discipline has to do with the way the latter seals off the 'economic' from other relevant aspects of the social. One important thread of the mainstream's interdisciplinary isolative strategy has been to assume the social and institutional system as given (and constant), and any mutual dynamic dependence between the 'economic' and 'social' realms irrelevant. For analytical purposes, individuals (and their tastes and preferences) are considered as given and the economy as a closed system.[15]

This is a system with no connections to its environment or to other systems: a circumscribed domain that is not influenced by external forces and does not exert influence on them. In a more detailed way, we may, following Dow and Chick,[16] define a closed theoretical system as one where all relevant variables are identifiable; the boundaries of the system are definite and immutable and consequently all

---

As Loasby (2005, 59) emphasised, "not only different sets of elements but also different ways of connecting a given set of elements define different systems".

[15] Of course, if pressed, virtually all economists would accept that the economic realm is an open system. However, the emphasis on becoming a hard science (some refer to 'physics envy') led to a central concern with mathematical tractability (particularly noticeable after the Second World War) and a consequent closing of economics.

[16] See Dow (2002, 139–140), Chick (2004, 6) and Chick and Dow (2005, 367).

variables can be clearly classified as exogenous or endogenous (these categories being fixed); the specified exogenous variables are the only ones that can impact on the system (which they do in a known way); the components of the system (variables, agents, and sub-systems) are separable (independent and atomistic) and constant in nature and the structure of their relationships knowable or random; the structural framework within which agents act, including all the information needed to determine the behaviour of the system, is assumed as given.

Even when some mainstream economists started crossing the traditional borders of economics to deal with more 'social' issues (such as crime, discrimination or family issues),[17] they did so without any significant change in their long-established rational-choice toolbox. Human behaviour continued to be analysed exclusively as equilibrium-efficient solutions for problems of allocation of known resources among alternative and clearly defined ends by given individuals. Institutional and other social mechanisms involved in the moulding of individual preferences and purposes remained ignored.

Although the work has been taken up with some (limited) success, there are many who find such asocial rational agents of economics wanting. An increasing number of scholars claim that such an approach does not actually take account of how the world works[18] and argue that an opening of our system of thought should be pursued.[19] Institutions and other social structures should be brought into the socio-economic system, not least because they have the capacity to reconstitute agents' preferences and purposes by acting upon ingrained habits of thought and action (cf. Hogdson 2002). The economy, according to this view, is not a self-contained system; it is an open subsystem of a much broader network – encompassing society, polity, and culture[20] – in the context of which it interacts with the other subsystems in a very complex way.

In an open system one or more of the conditions that Dow and Chick considered to be necessary for a closed system do not apply *in some degree*. Boundaries, instead of being isolative, are semi-permeable (allowing various in-fluxes as well as out-fluxes from other systems) and they are fuzzy and/or mutable. The constituent

---

[17] Since the 1950s.

[18] For instance, Wassily Leontief, the Nobel Memorial Prize laureate in Economics in 1973, wrote:
Page after page of professional economic journals are filled with mathematical formulas leading the reader from sets of more or less plausible but entirely arbitrary assumptions to precisely stated but irrelevant theoretical conclusions. ... Year after year economic theorists continue to produce scores of mathematical models and to explore in great detail their formal properties; and the econometricians fit algebraic functions of all possible shapes to essentially the same sets of data without being able to advance, in any perceptible way, a systematic understanding of the structure and the operations of a real economic system. (Leontief 1983: viii ff.)

The picture, unfortunately, remains up to date.

[19] The 'statements of principles' of some heterodox economics associations, for instance of the European Association for Evolutionary Political Economy: http://eaepe.org/node/5 and of the Society for the Advancement of Socio-Economics http://www.sase.org/about-sase/about-sase_fr_41.html are enlightening in this regard.

[20] And interlinked with the natural environment as well.

components of the system as well as the structure of their interrelationships may change in a non-predetermined way as they become in contact with other systems.[21]

Several are the voices who have argued for the need to open the discipline of economics and integrate social knowledge. It is not my purpose here to provide an historical account of the scholars and projects that were involved, but at least a brief mention should be made of K. William Kapp's (1961) work on a positive approach to the integration of social knowledge (a '*Science of Man in Society*'), arguably the first most elaborate modern proposal of this kind where the economy is thought of as an open system.

For Kapp, human actions and economic decisions did not take place in closed or even semi-closed systems; they occurred within a network of relationships involving four dynamic structures which were connected by a process of continuous open interaction: kinship, production and distribution, political systems and noetic systems[22] (cf. 1961, chapter VI). The economy was assumed to be an emergent component of a wider network of bio-physical-cultural relations, a stratum of a multi-layered reality.[23] In consequence, in his view, "a new approach which makes it possible to deal with the dynamic interrelations between economic systems and the whole network of physical and social systems and, indeed, the entire composite system of structural relationships" (1976, 97) was needed.

This would entail, Kapp thought, a search for an integration of social knowledge based on the definition of 'common-denominator concepts'[24] – understood as abstractions derived from observed regularities in behaviour (the so-called 'real types') – that would "provide an orderly and unambiguous framework for the comprehension of the elementary facts and uniformities of human nature and social processes" (Kapp 1961, 127). Those concepts should be comparable in comprehensiveness, generality and explanatory importance for social inquiry to

---

[21] von Bertalanffy, one of the originators of General System Theory, distinguished an open system from a closed one in terms of the interrelations with their respective environments. While a closed system was one where there is no exchange of materials, in an open system import and export would take place, with continuous influxes and outfluxes and a consequent change (generation and destruction) of the components of the system (cf. Bertalanffy 1973). Such a conception is here largely subscribed. However, following some more recent authors (cf., for instance, Chick and Dow 2005, Loasby 2003, Mearman 2005), I wish to emphasise that openness and closure *are not duals*; they occur in a continuum and we should think of them in terms of *dimensions* and *degrees* of closure.

[22] "Since these four substructures are connected by a process of continuous interaction, it follows that modifications in one must lead to transformation of the whole" (Kapp 1961, 115).

[23] A stratum (an entity or aspect) of reality is said to be emergent, or to possess emergent properties, "if there is a sense in which it has arisen out of some 'lower' level, being conditioned by and dependent upon, but not predictable from, the properties found at the lower level" (Lawson 1997, 176), thus turning the higher-level stratum irreducible to the lower-level one.

[24] "just as we look for a common denominator when we intend to add or subtract different fractions, so must we find or construct 'a set of common-denominator concepts in terms of which we can express *the otherwise incommensurable concepts* of our different disciplines, subject matters and cultures.' [quotation from Northrop]" (Kapp 1961, 127, emphasis added).

those of gravity in physics or evolution in biology. Accordingly, they would have to "cut across the subject matters of several disciplines" and "ought to be broad enough to encompass as many phases and aspects of human behavior as possible" (ibid, 128). For Kapp, *man* (and *human nature*) and *culture* were such concepts. Without integrating concepts, no communication between social disciplines would be possible.

In Kapp's view the "unified science of man in society" that would result from such an endeavour would respond to the "essential unity" and openness of the subject matter of the different social disciplines. However, it would not imply the unreasonable end of specialization and of division of labour in scientific work. This time specialization would be *problem-centred* (we could say thematic) and *determined by the nature of the problem or problem area under investigation* rather than disciplinary.[25]

It is enlightening to read K.W. Kapp in his own words:

> By making it clear that all social analysis deals with events and processes which occur within a broader and indeed unitary context, scholars working on special problems are forcefully reminded of the essential unity of their subject matter and come to see their special problems within the broader social context. In this way and not by imposition from the outside, lines of communication and real cross-fertilization will become possible and will actually take place. The traditional division of labor will then no longer block the way to interdisciplinary cooperation but will give rise to a tendency toward a general unification of scientific endeavors and results. In short, it is the unity of the subject matter and of the results of scientific research which will bring the social sciences closer together and induce social scientists to put their researches to the point where the present compartmentalization is replaced by a unified science of man in society. (1961, 204–5)

Kapp's project remains largely (and unfairly) ignored both in economics and in the social sciences more widely. Nevertheless his message has not been completely forgotten. Geoff Hodgson, for instance, one of the most distinguished contemporary European institutional economists, regrets what he called "the Grand Canyon in academic research and departmental organization between 'economics' and 'sociology'", responsible for the fact that "many of the most interesting questions in social science have become lost in the intervening abyss" (Hodgson 1994, 69), and projects economics "as just one facet of an eventually unified social science discourse, embracing sociology, anthropology, history, psychology and political analysis" (1996, 107).

More recently, Tony Lawson, the well-known Cambridge leader of the critical realist project of *economics as social theory*, similarly claimed that "there is no legitimate basis for distinguishing a *separate* science of economics". It would be "at

---

[25] This would mean to reject, as Kapp did, the current practice in economics of deducing conclusions from a narrow, disciplinary concept of man (*homo economicus*) which isolates one particular motif of human behaviour (calculative rationality) and leaves to other social sciences those ('non-rational') patterns of behaviour that do not take fully into account the real costs and expected benefits of one's actions. These other forms of behaviour, Kapp claimed, should also be considered as an integral part of our concept of man – of what he called the 'institutional man' (cf. Kapp 1968, 93)

most a division of labour within a single social science". Further, Lawson holds, "if economics is to be distinguished as a strand of social research, it cannot be according to its own ontology, methodological principles or substantive claims, but in terms of its particular focus of interest" (Lawson 2006, 499–500; see also 2003, 161–4).

Another important contribution towards an understanding of economics as an open system comes from the pluralist/Babylonian approach to knowledge, as developed by Sheila Dow and Victoria Chick in several of their works.[26] The two economists tackle the important and difficult issue of how to segment our object of study (conceived of as an open and organic totality) in order to make it manageable and to generate knowledge. In their view, a strategy of decomposing that object into isolated components/subsystems is legitimate and can be harmlessly pursued by means of *partial* analyses if the ultimate interconnectedness of things is kept at the back of our minds and is at some point brought to the fore. This means that such partial analyses must also be *provisional*. Further, they claim, as in the case of a rope, the strength of our studies increases when connections, within and across systems, are taken into account by employing a range of "parallel, intertwined, and *mutually reinforcing*" strands of thought and chains of reasoning.

12.5 Conclusion

The real world is an evolving and complex network of interconnected bio-physical-cultural open systems yet our possibilities of knowing it as a unity are rather constrained. The challenge is how to decompose our object of study (and build connections) in a way that makes co-ordination of those partial attempts to generate knowledge coherent. This, I believe, is the fundamental problem behind concerns regarding unity/disunity of science.

Unity/disunity and specialization/unification tend to be thought of in terms of duals. This misguides discussion. Trends towards both integration and disintegration characterize recent scientific developments. Specialization and attempts at unification are reasonable, although many times also conflicting, responses to the need to advance knowledge. Both have an ontological rationale. While specialization has its foundation in the human cognitive limits, the latter is to be associated to the interconnectedness of the world.

A sounder approach to the problem of unity/disunity of science, as suggested in this chapter, is to construe and develop sciences as theoretical open systems. We need open systems of knowledge in order to manage our cognitive limits and the ultimate interconnectedness of our world. In the social realm this entails conceiving and working towards an eventually unified social science, in the context of which each social discipline is merely a branch, justified by the unavoidable division of labour and distinguished from the other disciplines by its particular focus of interest.

---

[26] See, in particular, Dow (1996, chapter 2, and 2003) and Chick and Dow (2001).

This represents a major challenge for our current disciplines. Adopting an open systems approach will have deep implications on how each discipline is understood and developed. As we have seen, important work has already been done in this direction but a lot remains to be done in terms of clarification and elaboration. One fundamental task has to do with what constitutes the 'focus of interest' of each social discipline. In particular, what is the meaning of the 'economic' in this new context? How is economics differentiated from sociology and the other social disciplines? These are questions requiring a reply and further research.

References

Becker, Gary. 1993. Nobel lecture: The economic way of looking at behavior. *Journal of Political Economy* 101(3): 385–409.

Bertalanffy, Ludwig von. 1973. Théorie Générale des systèmes. French translation of *General System Theory* (1968). Paris: Dunod.

Chick, Victoria. 2004, Jan–Mar. On open systems. *Brazilian Journal of Political Economy* 24(1): 3–16.

Chick, Victoria, and Sheila Dow. 2001. Formalism, logic and reality: A Keynesian analysis. *Cambridge Journal of Economics* 25: 705–721.

Chick, Victoria, and Sheila Dow. 2005. The meaning of open systems. *Journal of Economic Methodology* 12(3): 363–381.

Coase, Ronald. 1994. *Essays on economics and economists*. Chicago/London: The University of Chicago Press.

Davis, John. 2006. The turn in economics: Neoclassical dominance to mainstream pluralism? *Journal of Institutional Economics* 2(1): 1–20.

Dow, Sheila. 1996. *The methodology of macroeconomic thought: A conceptual analysis of schools of thought in economics*. Cheltenham: Edward Elgar.

Dow, Sheila. 2002. *Economic methodology: An inquiry*. Oxford: Oxford University Press.

Dow, Sheila. 2003. Babylonian mode of though. In *The Elgar companion to post Keynesian economics*, ed. J.E. King, 11–14. Cheltenham: Edward Elgar.

Hacking, Ian. 1996. The disunities of the sciences. In *The disunity of science: Boundaries contexts and power*, ed. Peter Galison and David Stump, 37–74. Stanford: Stanford University Press.

Hogdson, Geoffrey. 1994. The return of institutional economics. In *The handbook of economic sociology*, ed. Neil Smelser and Richard Swedberg, 58–76. Princeton: Princeton University Press.

Hogdson, Geoffrey. 1996. Towards a worthwhile economics. In *Foundations of research in economics: How do economists do economics?* ed. Steven Medema and Warren Samuels, 103–121. Aldershot, Hants: Edward Elgar.

Hogdson, Geoffrey. 2002. Reconstitutive downward causation: Social structure and the development of individual agency. In *Intersubjectivity in economics: Agents and structures*, ed. Edward Fullbrook, 159–180. London/New York: Routledge.

Kapp, K.William. 1961. *Toward a science of man in society: A positive approach to the integration of social knowledge*. The Hague: Martinus Nijhoff.

Kapp, K. William. 1968. In defense of institutional economics. *Swedish Journal of Economics* 70: 1–18 (reprinted in Samuels, Warren (1988), *Institutional Economics I*, Aldershot, Hants: Edward Elgar, 92–107)

Kapp, K.William. 1976. The open-system character of the economy and its implications. In *Economics in the future: Towards a new paradigm*, ed. Kurt Dopfer et al. London: Macmillan Press.

Lawson, Tony. 1997. *Economics and reality*. London/New York: Routledge.

Lawson, Tony. 2003. *Reorienting economics*. London/New York: Routledge.

Lawson, Tony. 2006. The nature of heterodox economics. *Cambridge Journal of Economics* 30(4): 483–505.

Leontief, Wassily. 1983. Foreword. In *Why Economics is not yet a Science*, ed. Alfred S. Eichner, vii–xi. London: Macmillan Press (Originally published as a Letter to the Editor in *Science*, Vol. 217, July 9, 1982, 104–105).

Loasby, Brain. 1999. *Knowledge institutions and evolution in economics*. London/New York: Routledge.

Loasby, Brain. 2003. Closed models and open systems. *Journal of Economic Methodology* 10(3): 285–306.

Loasby, Brain. 2005. Making connections. *Econ Journal Watch* 2(1): 56–65.

Mäki, Uskali. 1992. On the method of isolation in economics. *Poznan Studies in the Philosophy of the Sciences and the Humanities* 26: 317–351.

Mäki, Uskali. 2001. The way the world works (www): Towards an ontology of theory choice. In *The economic world view: Studies in the ontology of economics*, ed. Uskali Mäki, 369–389. Cambridge: Cambridge University Press.

Mäki, Uskali. 2004. Theoretical isolation and explanatory progress: Transaction cost economics and the dynamics of dispute. *Cambridge Journal of Economics* 28: 319–346.

Mearman, Andrew. 2005. Sheila Dow's concept of dualism: Clarification, criticism and development. *Cambridge Journal of Economics* 29: 1–16.

Pombo, Olga. 2006. *Unidade da Ciência: Programas, Figuras e Metáforas*. Charneca da Caparica: Edições Duarte Reis.

Potts, Jason. 2000. The new evolutionary microeconomics: complexity, competence and adaptive behaviour. Cheltenham: Edward Elgar.

Runde, Jochen. 1996. Abstraction, idealisation and economic theory. In *Markets, unemployment and economic theory: Essays in honour of Geoff Harcourt*, vol. 2, ed. P. Arestis, G. Palma, and M. Sawyer, 16–29. London/New York: Routledge.

Chapter 13
Plurality of Science and Rational Integration of Knowledge

Catherine Laurent

Abstract Most scientists who are willing to articulate knowledge from different sources, for example to solve a problem set by the society, are confronted daily with the heterogeneity of scientific approaches and with the need to solve increasing difficulties in combining knowledge from various scientific areas. But these difficulties may be set aside when the philosophy of science postulates the possibility of establishing a priori rational principles of a unified science. In so doing it logically considers that the problems encountered in integrating scientific knowledge from different sources result from institutional gaps or organisational failures, or from the lack of researchers' skills. The analysis of these problems is consequently handed over to the sociology of science. Therefore, the debate on the unity of science may conceal the heuristic value of approaches based on an epistemological regionalism to support the rational construction of knowledge. Due to the heterogeneity of research programmes that coexist within each discipline the integration of scientific knowledge is an epistemic situation that needs clarification. Through short examples taken form social sciences and ecology, this chapter aims at showing how the formalisation of the plurality of scientific approaches may help to overcome difficulties encountered by researchers in getting an overview of the existing scientific knowledge, in integrating their approaches and in securing the recognition of the quality of their results.

Keywords Ecology • Economics • Interdisciplinarity • Pluralism

O.Neurath shared with R.Carnap and C.Morris the project to work towards a better integration of scientific knowledge. But he denied that the construction of an homogeneous system of laws for the whole science could be a relevant approach for

---

C. Laurent (✉)
INRA-SAD. UMR 1048, 16 rue Claude Bernard, 75 231 Paris cedex 5, France
e-mail: catherine.laurent@grignon.inra.fr

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology, and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_13,

articulating and using existing scientific knowledge at a given historical moment. "*The historical tendency of the unity of science movement is towards a unified science departmentalised into special sciences, and not towards a speculative juxtaposition of an autonomous philosophy and a group of scientific disciplines. If one rejects the idea of such a super science as well as the idea of a pseudo-rationalistic anticipation of the system of science, what is the maximum of scientific co-ordination which remains possible? The answer given by the unity of science movement is: an encyclopedia of unified science. (. . .) It may happen that one must use in one hypothesis, destined to a particular purpose a supposition which contradicts another supposition used in another hypothesis, destined for another particular purpose. One may try to eliminate such contradictions, but in the historically given science, and so in a real encyclopedia, these and other difficulties always appear. Encyclopedism may be regarded as a special attitude; one may also speak of encyclopedism as a program.(. . .) An encyclopedic integration of scientific statement, with all the discrepancies and difficulties which appear, is the maximum of integration which we can achieve.*" (Neurath 1938/1971; vol 1, p. 20).

Contradictions and discrepancies have not disappeared. Most scientists who are willing to articulate knowledge from different sources, for example to solve a problem set by the society, are confronted daily with the heterogeneity of scientific approaches and with the need to solve increasing difficulties in combining knowledge from various scientific areas. But these difficulties are ignored when the philosophy of science postulates the possibility of establishing a priori rational principles of a unified science. In so doing it logically considers that the problems encountered in integrating scientific knowledge from different sources result from institutional gaps or organisational failures, or from the lack of researchers' skills. The analysis of these problems is consequently handed over to the sociology of science. Therefore, the debate on the unity of science may conceal the heuristic value of approaches based on an epistemological regionalism to support the rational construction of knowledge.

Actually, for many scientists contributing to the on-going integration of scientific knowledge, the question of the ultimate state of science – unity *versus* disunity – does not matter so much. What is important is to acknowledge that here and now:

1. to deal with a similar phenomenon, theories based on distinct conceptual architectures – and sometimes contradictory hypotheses – coexist (for example Cartwright 1999; Mitchell 2002);
2. each research programme has its own area of demonstration and builds knowledge through specific paths (Bachelard 1949);
3. the social sciences, natural sciences and technical disciplines may produce knowledge with distinct properties which generates different kinds of interactions with its environment (Hottois 1996).

If these statements are to be considered seriously the integration of scientific knowledge is an epistemic situation that needs clarification to overcome difficulties encountered by researchers in integrating their approaches and in securing the recognition of the quality of their results. First of all, there is a necessity to get an overview of the existing scientific knowledge. One of the problem is that the exponential increase of scientific productions severely limits the interest of any encyclopedic programme. There is a need for new meta-knowledge that allows to better formalise the plurality of scientific approaches. This paper aims at showing how description of the sciences landscapes adapted from the notion of "research programme" from Lakatos and taking care of the properties of the different types of scientific knowledge may help this process.

13.1 Breaking Away from the Ambiguity of the Discipline

To observe science in progress in a perspective of epistemological regionalism, to acknowledge the plurality of sciences, we need to account for the way in which the different 'regional rationalisms' pointed out by Bachelard (1949) are constituted. The first idea that comes to mind is to use the concept of a discipline. This is not a very good idea.

Debates on the evolution of science generally refer, more or less explicitly, to a 'vulgate' that defines the discipline as[1]: '*… a set of rules and definitions, a domain of objects, a body of hypotheses considered to be true, theoretical and technical instruments, and a history (cf. M. Foucault). A discipline consists of a set of statements whose organization and coherence are regulated.' (Popelard and Vernan* 1997*: 20)* [our translation]

But this definition of discipline is misleading for several reasons, especially because: (i) a discipline is also (even primarily, according to Toulmin (1972)) a profession, with all its institutional dimensions; (ii) it is furthermore a material arrangement with its ad hoc instruments and material assets capitalized on over time (databases, model plants and model animals, long-term experimental devices, etc.), and (iii) a discipline does not necessarily correspond to a set of coherent statements; it usually groups together research programmes based on heterogeneous theories, which can have contradictory underlying hypotheses. This characteristic, described many times for the social sciences, also concerns the natural sciences, as N. Cartwright (1999) pointed out in the case of physics.

In other words, a discipline has to be defined as a complex set which necessarily comprises intellectual components (research programmes which can be based on contradictory hypotheses), institutional components (institutions that guarantee the validation and teaching of the knowledge produced) and material components (tools, databases, etc.). Its history (its historical background) has to be analysed from a threefold perspective since a discipline is based intellectually on a genealogy of problems, professionally on a series of institutional authorities, and materially on a set of technical devices (Laurent 2003).

---

[1] This definition is taken from a manual designed to popularize the philosophy of science, and the authors themselves repeatedly point out the incompleteness of a manual of this type.

The cooperation between disciplines brings into play these different dimensions. Consequently, even if there is a relative consensus to consider that the classification of the terms *pluridisciplinarity*, *interdisciplinarity* and *transdisciplinarity*, in that order, tends to signify a growing integration of disciplinary approaches, as O. Pombo (2004) noted there is clearly no agreement on what the object of integration is. It is considered that what counts is above all the degree of conceptual integration (Berthelot 2001) or the question of the sharing of a formalism between disciplines (Delattre), or else the questioning of the limits of scientific disciplines in all their dimensions (Stengers 2000) or, finally, the transcending of the discipline-profession, through partnership with laypersons to distinguish 'interdisciplinarity' from 'transdisciplinarity' (Gibbons et al. 1994).

This shows how much, when we analyse the question of the integration of knowledge in an internal perspective, the concept of a discipline can be a source of confusion rather than clarity. For example, what exactly does economy-ecology interdisciplinarity encompass? Which economic programmes (standard approach, economics of historical institutionalism, etc.) and ecological programmes (populations biology, landscape ecology, etc.) are they linked? To undertake an internal analysis, this indeterminacy compels us to depart from the discipline and to adopt a perspective set in conceptually coherent entities.

But then, which concept can be used to account for the heterogeneity of the world of scientific knowledge? K. Popper (1956) maintained that everything ought to be related to 'problems'. D. Andler (2002) suggested, however, that if Popper was able to sacrifice the discipline, it was *'perhaps because he shared with his opponents of the Vienna Circle a unitary conception of science: if basically science is one, why worry so much about the technical division of tasks within it, a division that evolves and affects neither the foundations nor the goals of the scientific enterprise? If at least he professes realism, the contemporary philosopher can rightfully hesitate; depriving himself completely of disciplines could lead him, without a unitary science, to give up science itself.' (p. 695)* [our translation]

Even if, in practice, those who study science in progress usually fall back on disciplines, there are alternatives. The concept of a 'paradigm' has been used in an attempt to account for the heterogeneity of scientific approaches over time (Kuhn 1962) or at a particular stage (Berthelot 2001). But actually, for the latter usage, the concept is rather fuzzy and provides little structure for undertaking an internal analysis, even a posteriori. The concept of a 'research programme', proposed by Lakatos (1970) seems more promising in so far as it offers a general framework to describe the different conceptual components of a research activity, irrespective of its institutional affiliation.

At a given point in time a research programme can be defined as a conceptual unit that coherently combines: (i) a permanent hard core of general theoretical hypotheses which are considered to be irrefutable by those in charge of the programme (e.g. hypotheses relative to the 'rationality' of human behaviours in the social sciences); (ii) "auxiliary hypotheses", ad hoc protective hypotheses, intended to protect this hard core and research under way from the 'anomalies' which they could encounter, that is, phenomena that are not consistent with the theory and that withstand observation (e.g. all the hypotheses that serve to define the domain of validity of results by adding a clause on everything that has to be covered by 'all things being equal'); and (iii) hypotheses to be tested, designed to expand the world of explainable facts. The concept of a research programme thus may serve to overcome several difficulties generated by that of a discipline for an internal analysis. It is a particularly useful tool for circulating within disciplines – and between them – and for comparing different approaches by clarifying the initial hypotheses on which they are based, by showing the approximations they entail, and by specifying the content of the ad hoc hypotheses that reduce the scope of their results.

We will now consider two examples to see how, by taking into account the heterogeneity of the hypotheses underlying different research programmes, we can understand certain internal components of the difficulties encountered by researchers in getting a clear vision of existing scientific knowledge, in articulating their researches and in securing recognition of the quality of their results.

13.2 Example 1: Coexistence of Research Programmes with Heterogeneous Premises

13.2.1 Mapping Science Landscapes

It is possible to describe on the same bases research programmes which are affiliated with the same discipline but are grounded in contradictory hypotheses. For example, two research programmes in economics can be compared that way as Fig. 13.1 shows. In this example we see that the hard cores of the two programmes under consideration[2] correspond to totally different domains of investigation and are based on contradictory hypotheses, for example as regards individuals' behaviour.

We thus observe that one case is an exploration of the heuristic value of a theory in terms of which individuals' behaviours stem from internal, conscious deliberation, whereas in the other case these behaviours are partially determined by sub-conscious processes resulting from mental structures that partly internalize social structures. In the former case the individuals' "irrational" behaviours, from an economic point of view, are anomalies against which the hard core has to be protected. This can be done by referring to constraints outside the field of economics. An attempt is made to eliminate the impact of these constraints by limiting the domain of validity of the model by means of ad hoc hypotheses. For example, we can consider that irrational behaviours have a biological determinant (for ex. lack

---

[2] Further information on these approaches can be found for neo-classical economics in Guerrien (2004) and Allen (1986), and for historical institutionalist approach in Boyer (1990) and Jessop (1997).

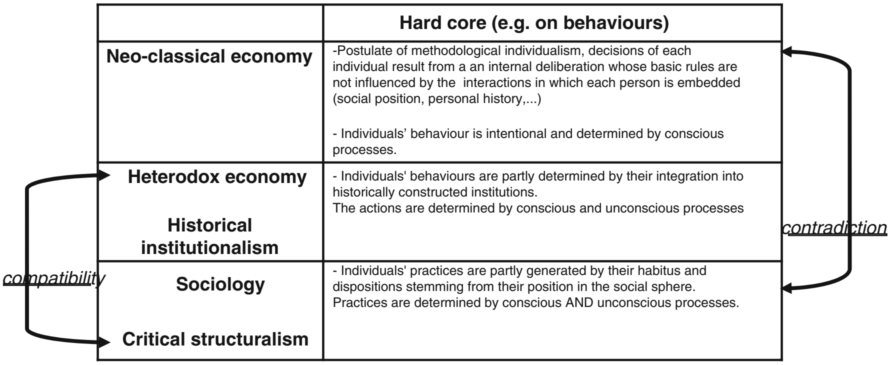
	Hard core (irrefutable hypotheses)	**examples of auxiliary hypotheses (e.g. on behaviours)**	**examples of fields of hypotheses to be tested on agriculture**
Neo-classical economics	- Institutions are entities defined by their functional role; the market, in which prices are set, plays a key part in individuals' socialization. - Harmonious economic functioning can be obtained in a context of perfect competition. *- Postulate of methodological individualism stating that the decisions of each individual result from a an internal deliberation whose basic rules (maximisation of utility...) are not influenced by the interactions in which each person is embedded (social position, personal history,...).*	Behaviours considered "irrational" may be explained by biological factors (e.g. procrastination for an individual who spends part of his income for "irrational" outlays such as inviting relatives instead of buying fertilizers)	- Configuration of agricultura markets in a framework of perfect competition. - Models of allocation of resources at the agricultural household level.
Heterodox economics **Historical institutionalism**	- The process of capital accumulation is decisive in an overall economic dynamic. - It is not spontaneously balanced by the market and competitive dynamic, and has varying forms in space and time. - Institutions and structural forms are decisive for channelling this process through a set of collective and individual behaviours. *- Individuals' behaviours are partly determined by their integration in historically constructed institutions.* *Actions are determined by conscious and unconscious processes.*	The emergence of new institutions is explained by the emergence of collective procedures during conflicts … which relate to sociology and to political science. They will be analysed ex-post.	**-** The roles of agriculture in the regulation of regimes of accumulation in different countries. - The evolution of institutional integration of different forms of agricultural activity.

Fig. 13.1 From the discipline to research programmes

of intellectual capacities), and construct auxiliary hypotheses accordingly. In the latter case, by contrast, the problematic aspect is the observation of new behaviours and especially the mechanisms through which they result in the creation of new institutions. Here again, auxiliary hypotheses allow for the protection of the hard core. Note, however, that in each case the range of hypotheses to test and the scientific facts constructed are directly related to the characteristics of the hard core and that the same designation (e.g. 'institution') can therefore encompass very different scientific facts.

Any research can be described through this notion of "research programme", even if the scientists concerned have not undertaken it with any formalisation of this kind, of the theoretical structure on which the research practices are based. Moreover, it does not stand to reason that in all circumstances this formalisation would be useful and more fruitful than the strategies of discovery which are implemented. The idea here is therefore not to use the notion proposed by Lakatos in a normative way, but only to point out that the concept of a 'research programme' can be used for description of the plurality of science, to construct internal analyses of science in progress, and to make an overview of sources of available knowledge on a subject and the hypotheses on which they are based.

This approach makes it possible to sketch the landscape of a discipline in a way that is accessible to those who are not directly part of it, and to the communities that keep this discipline alive. The possibility thus opened is not a trivial result. The history of scientific ideas is taught only in a limited number of syllabus. Even within a single discipline researchers usually have, outside their own field of investigation, limited knowledge of the heterogeneity of the approaches. The situation is worst for a scientist confronted with research programmes belonging to other disciplines. In that case, it is almost impossible to assess the precise consequences of the choices of hypotheses that are made in these diverse research programmes. Outside the



	Hard core (e.g. on behaviours)
Neo-classical economy	-Postulate of methodological individualism, decisions of each individual result from a an internal deliberation whose basic rules are not influenced by the interactions in which each person is embedded (social position, personal history,...) - Individuals' behaviour is intentional and determined by conscious processes.
Heterodox economy **Historical institutionalism**	- Individuals' behaviours are partly determined by their integration into historically constructed institutions. The actions are determined by conscious and unconscious processes
Sociology **Critical structuralism**	- Individuals' practices are partly generated by their habitus and dispositions stemming from their position in the social sphere. Practices are determined by conscious AND unconscious processes.

Fig. 13.2 The question of conceptual compatibility

scientific field, for people using this knowledge for practical purpose, the differences between research programmes are often analysed as subtleties that are too complex and difficult to grasp. That is why mapping science landscape is a priority to allow cross assessment of the domain of validity of the knowledge produced by the various theoretical frameworks that coexist.

13.2.2 Integrating Knowledge

13.2.2.1 Contradictions Between Research Hypotheses

By making it possible, on common bases, to re-examine research programmes that are institutionally affiliated to different disciplines, it becomes easier to identify and to take into account possible contradictions which can stem from the will to combine conceptual architectures based on incompatible hypotheses. If we pursue our reasoning on the basis of the preceding example, we see the full advantage, for researchers who wish to articulate sociological and economic approaches to a common object (individuals' behaviour), of investigating the compatibility of the hypotheses of the hard cores of the theories to which they are attached. Figure 13.2 provides an example of this approach.

In the first research programme, the forms of individuals' social embedness can be neglected in order to apprehend their behaviour. By contrast, in the second and third research programmes[3] individuals have to be characterized, at least partially, by the forms of their social position which contribute towards guiding their behaviours. Consequently, it seems coherent to try to connect the observation frameworks and results of the latter two research programmes while the articulation

---

[3] Further information on this sociological theory can be found in Bourdieu, Passeron 1989.

of an economics based on a postulate of methodological individualism, to a sociology of critical structuralism seems to be the source of many logical difficulties. These forms of complementarity or incompatibility can be analysed in this way between different disciplines.

13.2.2.2 Unequal Impact of Questions Set by the Society on Research Programmes Heuristic

The advantage of the concept of a research programme is that it also allows for observation of the way in which approaches will be revised to ensure the programme's progress or, under 'external' constraints, to try to produce useful knowledge for answering a question raised by society. In the latter case, we observe that the hypotheses to test, derived from questions asked by society, will have heterogeneous statuses, depending on the research programme. For instance, to study the conservation of biodiversity in agricultural landscapes (Fig. 13.3), it is necessary to test hypotheses relative to interactions between ecological and social processes. But the importance of these hypotheses will differ as regards the advancement of the positive heuristics of the research programmes concerned. This is a key question for landscape ecology,[4] but one that is entirely secondary for economics or even one that should momentarily be excluded to safeguard the positive heuristics of the programme.

In addition, to deal with spatial ecological processes, approaches in economics (whatever their theoretical background) will have to collect data with spatial attribute (and design samples accordingly) which may result into an enormous supplementary load of work when compared with the classical procedures of the discipline. Enormous load of work for dealing with a question that is marginal or the positive heuristic of the research programmes in economics . . .

13.2.3 Assessing Research Results

We thus observe that by compelling certain programmes to reconsider sections of the reality that they did not include in their theoretical concerns, the collaboration between research programmes imposed by the necessity of practice can be unequally disruptive for the programmes concerned.

---

[4] Further information on this approach can be found in Burel, Baudry 2003.

	Hard core *(irrefutable hypotheses)*	***Fields of hypotheses to test so that the original programme can progress***	***Hypotheses to test to answer a question raised from the outside***
Standard economics	- Institutions are entities defined by their functional role; the market, in which prices are set, plays a key part in individuals' socialization - Harmonious economic functioning can be obtained in a context of perfect competition. - Postulate of methodological individualism stating that the decisions of each individual result from a an internal deliberation whose basic rules (maximisation of utility...) are no influenced by the interactions in which each person is embedded (social position, personal history...).	- Configuration of agricultural markets in a context of perfect competition. - Models of resource allocation at the level of exploitation or of the farm household.	*Existence of causal relations between economic dynamics, of agriculture and ecological processes.* ***Methodological consequences*** *Necessity of integrating the spatial dimension for both the collect and the processing of data.* *Necessity of considering all agricultural activities with significant impact on landscape functioning irrespective of their economic importance (e.g.. to integrate small scale farms with little connection to the market).*
Heterodox economics Historical institutionalism	- The process of capital accumulation is decisive in an overall economic dynamic. - It is not spontaneously balanced by the market and competitive dynamic, and has varying forms in space and time. - Institutions and structural forms are decisive for channelling this process through a set of collective and individual behaviours. - Individuals' behaviours are partly determined by their integration in historically constructed institutions. Actions are determined by conscious and unconscious processes.	- Roles of agriculture in the regulation of regimes of accumulation in different countries. - Evolution of institutional integration of different forms of agricultural activity.	
Landscape ecology	- Landscapes are heterogeneous ecological systems that are hierarchically structured in space and time. Their behaviour is driven by both internal and external (spatial interactions) factors. - Even if local transient states reach equilibrium, the global heterogeneity precludes global equilibrium. -Human are part of the system and their activities are ecological factors.	- Spatial and temporal dynamics of ecological processes in anthropised environments.	

Fig. 13.3 The hypotheses to test have heterogeneous statuses, depending on whether they have been formulated by researchers or stem from a request by practitioners

13.3 Example 2: Fundamental Otherness of Different Types of Scientific Knowledge

13.3.1 Mapping Science Landscapes

We observe that not all research programmes produce statements with the same properties. Hacking (1999) highlighted the particular aptitude of the social sciences to produce knowledge of an interactive nature, and to produce classifications that induce a dynamic interaction with the classified individuals as soon as they are aware of that classification and their place in it. By contrast, the natural sciences primarily produce knowledge without any direct effect on the behaviour of the classified entities.

This question of the specificity of the social sciences stemming from the adjustment of human behaviours to the description of a situation is not new (e.g. Neurath 1939 p. 26 et suiv.). In a famous article on self-fulfilling prophecy, sociologist R. Merton (1949) considers behaviours which make a situation real because they are based on the belief that it is real. For instance, the customers of a bank cause its bankruptcy by withdrawing all their money because they believe that it is going bankrupt, when in fact it is financially healthy. In the examples chosen, the description of the initial situation is false and is based on rumour. But when it is known by the people concerned, the validated scientific knowledge can trigger adjustments in behaviours and thereby alter the characteristics of the observed situation.

These observations concern the human sciences above all, but there is no strict correspondence between types of knowledge and types of science. Genetics, for example, produces knowledge of an interactive kind when it causes individuals to be classified in terms of their risk of developing certain pathologies, and when at-risk subjects are informed of this classification. This can be a source of anxiety which aggravates the risk or, alternatively, may prompt the person concerned to adopt preventive measures that reduce the risk. Various disciplines can thus produce knowledge of an indifferent or interactive kind, even if one of these forms of knowledge may be dominant in a particular scientific field.

We might add, following G. Hottois (1996), that in addition to the symbolic interactions mentioned above, knowledge can also be a source of immaterial or material technical interactions when it leads to the creation of tools or the transformation of its material environment.

The production of knowledge of an interactive kind thus creates situations in which relations between scientific statements and prediction are totally different from those in which research produces knowledge of an indifferent kind. The consequences of these interactions reduce the predictive capacities of theories because even if at moment t a classification can make certain behaviours intelligible, at $t + 1$ the classified individuals who have taken this classification into account may change their behaviour in various ways. It is therefore possible to have statements or self-fulfilling theories, that is, theories which will be proved provided that they are considered to be true by the actors concerned, and those actors adjust their behaviours accordingly. We can show, for example, that a theory relative to certain economic fluctuations will be verified if each agent acts in the belief that the theory is true (Chiappori et Guesnerie 1988). But a statement can also induce its own invalidation if the classification is associated with a strong probability of an event judged unfavourable, and the person classified has modified his or her behaviour to avoid that outcome. This makes it necessary to consider the principles of production of evidence from a specific perspective, for the question of what is true and what can be proved becomes highly problematical in the case of knowledge of an interactive nature. It also makes it necessary to examine the place granted to the quality of predictions, if the results of a scientific approach are to be judged. This is a permanent source of heterogeneity of procedures for validating rational knowledge since, depending on the kind of knowledge produced, the validation of results demands the implementation of different procedures.

13.3.2 Integrating Knowledge

A question arises when concrete research practices aim at integrating knowledge from sources: of what nature is the knowledge that will be produced when a model combines variables of different kinds?

The analysis of precise examples shows that when a model is mixed and combines variables of an indifferent kind (e.g. ecological classification of the richness of the flora in certain areas) and variables of an interactive kind (e.g. typology of farmers' practices), it will produce classifications whose properties are those of the classifications of an interactive kind. Hence, it is possible to produce typologies that combine these different sorts of variable, in order to link grazing practices to richness of flora. When they are known to the people implementing the practices concerned, such typologies can modify their future behaviour. Assuming that ecologists recognize and disseminate information claiming that a given grazing practice has a positive impact on the richness of the flora, this practice may then be adopted or rejected by farmers for various reasons: technical (a way of considering the quality of the grazing), symbolic (the farmer's personal interest in or rejection of certain arguments defending the environment), economic (anticipation or not of a possible remuneration linked to the maintenance of certain plant species), etc. All in all, the diffusion of this kind of typology has new effects on plant-related practices and dynamics which are not predictable within the limits of the initial model.

13.3.3 Assessing Research Results

This example shows that in certain circumstances of knowledge integration, knowledge will fall into an interactive kind thus has fewer predictive qualities. Predictivity implies, among other conditions, an adequate stabilization of the object under study so that the knowledge of a law characteristic of one of its states can inform on its future behaviour (Godard and Legay 1992). But here, the classified individuals' knowledge of the highlighted regularity (on the interactions between practices and richness of flora) causes the destabilization of the object under study (dynamic of the richness of flora).

Thus, when they are engaged in pluridisciplinary models with the social sciences, researchers in the natural sciences can produce models whose predictive capacities are smaller than those usually required in their discipline. With a more complete approach to reality, the scientists concerned may be suspected of losing in scientificity from the strict point of view of the natural sciences where 'excellence' is often related to a certain quality of prediction.

13.4 Conclusion

The examples above show that it is possible to formalize shared principles of analysis which enable one both to take into account contradictions between research programmes and to construct a global vision of the regions of science. Such representations become indispensable for apprehending the internal difficulties encountered by scientists in rationally bridging the gaps between scientific communities and in specifying the conditions in which knowledge can effectively be incorporated in order to treat complex problems set by the society.

In a world in which the unity of science in progress were not problematical, the boundaries between scientific sub-communities would be easy to overcome. These sub-communities would not only share this ideal of rationality but would also aim to have compatible underlying hypotheses, and would have a common way of grading the procedures of revision of research programmes and of judging the results of investigations. But we are not in that world. From one research programme to the next, researchers' objectives differ. Moreover, the lack of knowledge about others' research programmes is growing. Interactions between sub-communities associated with competing research programmes, or programmes with very different research topics, are sometimes reduced to virtually nothing.

Yet, if as a last resort the validation of scientific statements is the outcome of critical intersubjectivity of a community (Fagot-Largeault 2002) that shares a demand for rationality, then the exercise of that critique is fundamental. Any renunciation of incursions into the knowledge of other communities, and any refusal of such incursions into that of one's own community, appear singularly threatening for the production of rational knowledge. This renunciation is not only the doing of researchers who refuse to look beyond the boundaries of their own research field; it can also take the form of a denial of the persistent heterogeneity of the conceptual bases of science in progress and, on the whole, this heterogeneity is seldom analysed. The three communities which, *a priori*, would have good reasons to feel concerned, avoid the problem. Scientists, who generally very legitimately reject theoretical eclecticism, focus on their own research programme since they have chosen the approach that seemed to them the best. Philosophers of science, who cannot in one movement embrace the growing complexity of the scientific scene, generally favour the dominant approaches in the social field of science, or confine themselves to *ex post* analyses, leaving it up to history to sort through rival paradigms. Research managers, harassed by this profusion and all the contradictions it entails, often prefer to assume that there are differences only in details, which can be ignored.

But, for a cross cutting critique to be possible, and for real progress to be made in the integration of scientific knowledge, it is necessary to build analytical frameworks that allow take into account that every research programme claims its own domain of demonstration and its own validation criteria. That is why it seems necessary, at a meta level, to construct analyses which on common bases explicate the hypotheses of heterogeneous research programmes, their criteria of scientificity, the part of arbitrary of each approach, and the limits of empirical validity of their results.

References

Allen B. 1986. General equilibrium with rational expectations. In *Essays and honours of Gerard Debreu*. Amsterdam: North Holland.

Andler, D. 2002. L'ordre humain. In *Philosophie des sciences*, vol. II, ed. D. Andler, A. Fagot-Largeault, and B. Saint-Sernin, 673–824. Paris: Gallimard.

Bachelard, G. 1949/1975. *Le rationalisme appliqué*, 215 p. Paris: PUF.

Berthelot, J.-M. 2001. *Epistémologie des sciences sociales.* Paris: PUF

Bourdieu, P., and J.-C. Passeron. 1989. *La reproduction*. Paris: Ed Minuit.

Boyer, R. 1990. *Regulation theory: A critical introduction*. New York: Columbia University Press.

Burel, F., and Baudry, J. 2003. *Landscape ecology, concepts, methods and applications.* 355 p. New Hampshire: Science Publishers INC.

Cartwright, N. 1999. *The dappled world. A study of the boundaries of science*, 247 p. Cambridge: Cambridge University Press.

Chiappori, P.-A., and R. Guesnerie. 1988. Endogenous fluctuations under rational expectations. *European Economic Review* 32(1988): 389–39.

Delattre, P. *Interdisciplinaires* (*recherches*), Encyclopaedia Universalis. Version 8. 2002. CDRom.

Fagot-Largeault, A. 2002. La construction intersubjective de l'objectivité scientifique. In *Philosophie des sciences*, vol. 1, ed. Fagot-Largeault Andler and B. Saint-Sernin, 129–225. Paris: Gallimard.

Gibbons, M., C. Limoges, H. Nowotny, S. Schwartzman, P. Scott, and M. Trow. 1994. *The new production of knowledge. The dynamics of science and research in contemporary societies*. London: Sage.

Godard, O., and J.-M. Legay. 1992. Modélisation et simulation: une approche de la prédictivité. In *Sciences de la nature. Sciences de la société. Les passeurs de frontières*, pt. 491–518. Paris: CNRS Editions.

Guerrien B. 2004. *La théorie économique néoclassique.* Paris: La découverte

Hacking, I. 1999. *The social construction of what?* Cambridge: Harvard University Press.

Hottois, G. 1996. *Entre symboles et technosciences*, 268 p. Paris: Champ Vallon – PUF.

Jessop, B. 1997. The regulation approach. *The Journal of political Philosophy* 5(3): 287–326.

Kuhn, T., 1962/1983. *La structure des révolutions scientifiques*, 284 p. Paris: Champs Flammarion.

Lakatos, I. 1970. History of Science and its Rational Reconstructions. In *PSA 1970 Boston Studies in the Philosophie of Science*, vol. 8, ed. R.C. Buck and R.S. Cohen, 91–135. Dordrecht: Reidel.

Laurent C. 2003. *Pour une épistémologie de l'interdisciplinarité*, 64 p. Paris: Université Paris I, INRA.

Merton, R.K. 1968/1949. *Social theory and social structure*, 702 p. New York: The Free Press.

Mitchell, S. 2002. Integrative pluralism. *Biology and Philosophy* 17: 55–70.

Neurath, O. 1939/1970. Foundations of the social sciences. In *Foundations of the unity of science. Towards an international encyclopedia of unified science*, vol. II, ed. O. Neurath, R. Carnap, and C. Morris, 1–51. Chicago/London: The University of Chicago Press.

Neurath O. 1938/1971. Unified Science as an encyclopedic integration. In *Foundations of the unity of science. Towards an international encyclopedia of unified science. International encyclopedia of unified sciences*, vol I, 1st ed. 1938, ed. O. Neurath, R. Carnap, W. Morris (Dir). Chicago: University of Chicago Press.

Pombo, O. 2004. *Interdisciplinaridade: ambições e limites*, 203 p. Liboa: Relogio d'agua.

Popelard, M.-D., and Vernan, D. 1997. *Les grands courants de la philosophie des sciences*, 95 p. Paris: Seuil.

Popper, K. 1956. On the non existence of scientific method. Preface for the first English edition. In *The Logic of scientific discovery*, 5–8. London: Hutchinson.

Stengers, I. 2000. Entretiens. Discipline et interdiscipline: la philosophe de "l'écologie des pratiques" interrogée. *Natures, Sciences, Société* 8(3–4): 51-58–59-63.

Toulmin, S. 1972. *Human understanding. The collective use and evolution of concepts*, 520 p. Princeton: Princeton University Press.

Chapter 14
A Physicalist Reconstruction of a Theory: The Case of the Freudian Theory of Hysteria

César Lorenzano

Abstract Dora, an 18-years young girl, goes with her father to Freud's clinic, on account of several close to disabling signs and symptoms: "*tussi nervosa*", aphonia, appendicitis attacks, etc. that Freud diagnoses as hysteria. During the treatment, he discovers that Dora has full knowledge of human sexuality, even of so-called perversions. She is in love with Mr. K, with Mrs. K, probably with her own father, she has intense relationships with other females, and she dreams with houses that catch fire. While retelling their interviews, Freud builds his psychoanalytical conception of hysteria. From that material – and following the distinctions of the structuralistic conception – I extract its components and fundamental laws, using them to describe Dora and the other cases of hysteria that resemble this paradigmatic case. This reconstruction of the *Freudian theory of hysteria* is carried without using – or starting with – the description of abstract mathematical structures, as usually happens in structuralism. Briefly, it is a reconstruction of the *applications* of the theory that complies with all the usual requirements of a structural analysis. In this way I solve some ontological tensions of this conception, showing that it can be developed with an ontological and epistemological monism proper of a physicalistic and nominalistic approach.

Keywords Freud • Hysteria • Nominalism • Physicalism • Psychoanalysis • Structuralist conception of science

---

C. Lorenzano (✉)
National University of Tres de Febrero, Rivadavia 3033 4° 34, Buenos Aires (1203), Argentina
e-mail: clorenzano@gmail.com

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology, and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_14,

14.1 Introduction

In 1905 Freud publishes the article *Fragment of an Analysis of a Case of Hysteria* – commonly call the *Dora's Case*[1]– written in two weeks five years after December 31 1900 when the patient herself put an end to her treatment.

As it lasted less than 3 years, Freud could remember it completely and register it as a clinical history, the canonical form in which the doctors communicate their cases.

In this paper I will summarize the main notions that Freud introduces in his article, and the relationships between them. This characterization will be presented by means of an informal reconstruction that conserves the basic tenets of the structuralistic conception of the theories, assuming that everything I say can be stated in a formal language without losing any content in the process.

Freud has not always followed in his text all the notes that he wrote during the treatment. Notwithstanding that, we can still witness a process in which certain facts are present, the interpretation he makes by means of a theoretical apparatus, as well as the ways he perceives new data corroborates the interpretation. We are spectators, thanks to his narrative style, of the genesis of an important portion of the psychoanalytical theory; in fact, of a paradigmatic case of his conception of hysteria that is at the same time the central, paradigmatic nucleus, of the other amplifications of such theory.

In the course of Dora's analysis, we see how empirical data force Freud to formulate theoretical presumptions, and to apply those already established, in such a way that the contexts of discovery and of justification of the fundamental tenets of psychoanalysis sometimes overlap or follow each other in a continues way.[2]

My reconstruction, therefore, will be as much of what happens to Dora, as of the Freudian conception of hysteria. The Dora's Case enacts the terminological apparatus and laws of psychoanalysis. As we will see further down, I will characterize hysteria by its applications, and not by its abstract models.

---

[1] Freud, S. (2000) Vol. VII (1901–1905) "Fragment of an Analysis of a Case of Hysteria (1901 [1905]) pp. 7–125. German Edition: Bruchstücke Einer Hysterie-Analyse, Mschr. *Psychat. Neurol.*, 18 (4 and 5) Oct and Nov., 285–310 and 408–467.

[2] Speaking of his scientific method, and of the discovering of the unconscious, Freud says (2000), Vol. VII, pp. 112–113: "I can only assure the reader that I approached the study of the phenomena revealed by observation of the psychoneurosis without being pledged to any particular psychological system, and that I then proceeded to adjust my views until they seemed adapted for giving an account of the collection of facts which had been observed." He adds: "I take no pride in having avoided speculation; the material for my hypothesis was collected by the most extensive and laborious series of observations." As he knew very well the standards of empiricism, he expresses: "But of this I am certain –that any one who sets out to investigate the same region of phenomena and employs the same method will find himself compelled to take up the same positions, however much philosophers may expostulate."

We know that it is after writing his clinical cases when Freud presents his psychoanalytical conception theoretically, with few empirical references.[3]

If we proceeded to analyze his thought from these later – rather abstract-articles, and tried to begin the reconstruction in the traditional way, characterizing the abstract models of the theoretical core, we would be ignoring the enormous empirical load of his theory. That is what makes it a factual conception, and not merely a theoretical one.

When I reconstruct hysteria by its paradigmatic application, namely Dora's Case, I am consequent with Freud himself, and with the medical thinking from which he comes. In general, there are no illnesses at all, just sick people; similarly, I will say that there is no hysteria, but hysterical people. There is nothing but cases of illnesses – and of hysteria – and their knowledge by doctors and psychoanalysts. However, text books simplify the knowledge of those individual, unique cases, eliminating their more particular aspects, and we only read their more general profiles. In the process it becomes *general* knowledge that is compatible with other similar cases. When applied, a new clinical history is written, and the inverse process takes place, endowing the new case with the whole specificity that it possesses.

On the other hand, as is well known by structuralism – following Wittgenstein – without paradigmatic cases we would not know how to use those purely theoretical statements, and we would not realize that they are immersed in the factual world from the very beginning.

The proposal of reconstructing the theory by its applications – and not by any abstract structures – is due to the conviction that factual knowledge could never be characterized correctly by non-interpreted mathematical or logical structures, and even when we need to use mathematical structures, they are already interpreted.[4]

As I mentioned before, this strategy introduces several modifications to the standard reconstruction of theories. The first one is to reconstruct the theory using the steps of the empirical claim,[5] as a natural and pedagogical way to do it. It recognizes its historical importance, since the empirical claim and its problems are in the origin of structuralism. The other modification is to use diagrams to show elements and relations of the theory instead of a formal notation. They are simpler than the mathematical symbols and illustrate a different way to represent the

---

[3] I refer to writing as: Freud (2000) Vol. XIV, "A History of Psycho-Analytic Movement", "Papers on Metapsychology and Other Works" (1914–1916), and Vol. XIX, "The Ego and the Id and Other Works" (1923–1925).

[4] I will use *abstract* in two senses. The first one refers *abstract* to non-spatiotemporal entities. The other one refers to non-interpreted mathematical or logical structures. I will try to show that both of them are superfluous in the foundations of empirical knowledge. The first sense is rejected by physicalism as pure platonic metaphysics. Of course, you have to be a physicalist so as to say so. I will try to argue –and show in my reconstruction- that to reconstruct scientific knowledge by means of logic or mathematical procedures all is needed –and all it may be used- are *interpreted* logic and mathematics, and not pure mathematics and logic, avoiding the discussion about the Platonic nature of its objects that is not the purpose of this article.

[5] See below: **The empirical claim as a strategy of reconstruction**.

structure of the theory. These pedagogical and historical reasons are good enough to justify the reconstruction of the *Freudian theory of hysteria* by means of applications and diagrams, and it could be adopted by the structuralistic community within the usual framework. It is more controversial to use them to solve an ontological and epistemological tension between a mathematical core with no interpretation, and a set of empirical applications as its Wittgensteinian semantics. There is a tension – and perhaps a contradiction – between a Platonist core and a nominalist use of the theory. I expect that the reconstruction that we develop further below will be considered appropriate to solve that tension in a homogeneous physicalist and nominalist way.

In what follows, I will present successively:

i. a synthesis of Dora's clinical history,
ii. a reconstruction of Freud's hysteria theory, in an amended version of the structuralistic conception;
iii. I will argue meanwhile about the physicalist consequences to work out a reconstruction of applications.

14.2 Fragment of an Analysis of a Case of Hysteria

"Fragment of an Analysis of a Case of Hysteria" is the original title of Freud's text, perhaps better known as Dora's Case. We will here present a synthesis, necessarily incomplete, just to offer the reader an overview of that history, and of Freud's procedures.

The patient – who is call Dora[6] – is an 18 year-old girl that is taken to Freud's clinic by her father, a man around 45 years of age, because she presents several signs that Freud unequivocally attributes to hysteria, a pathological entity well studied and typified by the medical science of the time. Just as in medical studies, Freud uses the word "signs" to talk about Dora's alterations. They are the following: "*tussi nervosa*", aphonia, migraine, appendicitis, depression of spirit, enhanced excitability, suicide warning leaving the letter in sight, *taedium vitae*.[7]

According to Freud, one of the signs that allowed him to identify hysteria is that the history told by those suffering the illness is not exact or coherent, and it possesses dark areas, even amnesias and voids.

By that time, Freud had already published *Studies on hysteria*[8] in collaboration with J. Breuer in which he established that the genesis of the illness includes traumatic facts and conflict of affections, and that different aspects of sexuality are

---

[6] Dora is the nickname that Freud uses for Ida Bauer, sister of Otto Bauer, the well known Vienna Marxist, who was related to Otto Neurath.

[7] These are medical terms used by the English translation of the paper.

[8] Freud, S. (2000) Vol. II (1893–1895) "Studies on Hysteria".

involved. Those events are hidden – they are repressed – and therefore they become pathogen. He had also published *The Interpretation of Dreams*,[9] in which he claims that those repressed events appear, transfigured, in the dreams.

There was an episode narrated by Dora, that her father, though not yet believing that it was true, judges as responsible for her depression of spirit, excitability and suicide notions.

The event that Dora had communicated to her parents consisted in the fact that in a walk with Mr. K, her father's friend, he made love proposals. In principle, this was the characteristic traumatic event stated in the *Studies on Hysteria*; but the ulterior facts discovered in Dora's analysis, namely that some of the signs had often appeared long time before this episode, even when she was eight, forces Freud to go beyond this first theory. This makes him look back in time. He found, indeed, that she had forgotten a kiss forced by K when Dora was 14 years old, and that causes her a sensation of repulsion, and other events that could be related to her childhood signs.

Freud concludes that in those times she was already hysterical, because if she was in love with K, it would have been normal to feel some genital excitement, but not repulsion. He believes that she noticed the erect member of K, and that this sensation was displaced upward, towards the respiratory system, causing "*tussi nervosa*" and hoarseness, as well as repulsion and nausea.

Dora's family did not pay any attention to the fact that K, who was older than Dora, sent flowers to her every day, and gave her important gifts.

Dora thinks that was so because her father, who was the lover of K's woman, consented to it so as not to disturb this relationship; in a certain sense, he gives her up to K.

Freud believes that even before this there were some traumatic episodes, because she showed symptoms when she was 8 years old. As we will see afterwards, they were revealed during the analysis of Dora's first dream.

Freud enunciates the rule that the symptom represents the realization of a fantasy of sexual origin.

As an example of his long interpretations, we will relate one of them. It begins when Dora says that her father is a man of (monetary) resources. Freud thought that its meaning was literally the opposite. In fact, she thought that her father had limited resources, another way of expressing that he was impotent; and that presumption was not contradictory with the relationships that he maintained with Mrs. K, because Dora suspects that they satisfied each other by means of oral sex. Freud stated that this was the probable origin of the signs that she experiences in her mouth and throat.

The fact that the cough disappeared after this explanation confirms the validity of the interpretation.

For those that believe that a young girl cannot have these thoughts and knowledge, Freud adduces that perversions have their natural origin in primary sexuality, and that all of us can surpass the limits of normality, although it usually is

---

[9] Freud, S. (2000) Vol. IV and V (1900–1901) "Interpretation of Dreams", "On Dreams".

sublimated, i.e. it is energy that is used for other ends, namely to produce culture. Freud thinks that the symbolic mechanism of associating images allows us to go from the erotic sensation in the lips during the suction of the breast by the infant – an example of primary sexuality-, to the penis, following an intermediate road that goes through the vision of naked babies, and of calves suckling from their mother.

Freud had already sketched the theory that the early love that babies feel for their parents – Oedipus complex – fixes the loving impulse when it takes the form of sexual inclination during puberty.

He relates us that Dora took care of her father during his periods of illness, lung tuberculosis, that forced him to live in special places, where she met K and his wife. She was jealous of her father, but she also felt homosexual jealousies for K's wife. The predominant, obsessive idea of her father's illicit relationships with Mrs. K was hiding in a completely unconscious way her love for K, and his wife. Freud thinks that when the sexual libido of a hysterical woman focused on a man is repressed, the homosexual tendency is intensified, and it can even become conscious. He says that contradictory ideas can coexist, as her love toward K, and toward his wife, even when the latter criticized him all the time.

Dora relates two dreams during her treatment. We will only narrate the first one. It is highly symptomatic of Freudian interpretations.

First dream: "A house was on fire. My father was standing beside my bed and woke me up. I dressed quickly. Mother wanted to stop and save her jewel-case. But father said: I refuse to let myself and the two children be burnt for the sake of your jewel-case. We hurried downstairs and as soon as I was outside I woke up."[10]

As it happens in some occasions, the dream is built with events that happen in the vigil. They are the following ones: The mother closed the dining room, and in doing so she also closed the exit of the brother's bedroom. The father arrives in the town where the family was in the middle of a storm, and he expresses the fear that lightening could cause a fire in the wooden house where they were vacationing.

Some of the elements that appear in the dream, as the fire, the closed bedroom that could damage the children, the father's protective attitude, were already there, in the events of that day.

When Freud goes deeper into the dream, Dora relates that she dreamed it the very day when K tried to kiss her: therefore, it is a kind of answer to that event. She remembers that K was near her while she slept in her bedroom; that's why she wants to close it with a key, but the key is not in its place (she thinks that K took it out so as to avoid that possibility). For that reason, she always gets dressed in a hurry (as she says in the dream). That dream lasted four nights, as long as they stayed in the vacation place.

Here the dream goes from an obvious daytime event, to another one connected to Dora's intimacy. Freud had already established previously that dreams are the representation of desires. When he goes deeper into it new material emerges that reinterprets the dream, and leads to other discoveries from Dora's past.

---

[10] Op. cit. Vol. VII, p. 64.

K had given her a case. It is a figure that makes allusion to the feminine genital. My case – my sex – is in danger, she says in the dream. Her father saved her. The mother appears in the dream, although she was not with them, because Dora is willing to give to her father what the mother denies him (the case). She had said before that she believed that her parents didn't have sexual relations. The mother's feminine figure can substitute Mrs. K who fakes to be sick to avoid having relations with her husband; so, in the dream Dora expresses now with regard to K, with the ambivalence that characterizes dreams, that she can give him what her wife denies to him. Freud interprets that Dora was afraid to surrender to her desires for K.

Here the infantile element appears for the first time in the interpretation: a dream has as an origin in a current event and also in a childhood one. The desire that the dream satisfies – the Oedipus complex – is infantile.

While they speak about the dream, Dora puts a box of matches on the table. Freud associates matches with playing with fire, and immediately with urinating in bed (an old saying states that he who plays with fire will wet the bed). Dora remembers something she had forgotten until that moment: she had suffered from bed-wetting until her 8 years of age. And bedwetting is caused always, according to common knowledge of those times, by masturbation that can also produce genital flow – leucorrhoea-. Dora admits she masturbates herself. She knows that her father had venereal illnesses and she suspects he infected her mother, and that was why she had a flow. When Freud tells her so, she plays "with a small reticule of a shape which had just come into fashion, opening it, putting a finger into it, shutting it again, and so on", a symptomatic act whereby she acknowledges that the interpretation was correct.

In the long interpretation of the dream some repressed memories arise that have to do with hysteria episodes that Dora had when she was 8 and 14 years old.

Signs of her hysteria such as dyspnoea, asthma, are imitations of her father when he snorts during the coitus, a memory that had been repressed. A similar imitation of father's cough is the cause of her "*tussi nervosa*".

The desire to replace K by her father provides the energy that causes the dream.

The scent of smoke she remembers in the dream has to do with the smell of tobacco in K's breath when he kissed her.

Here I stop telling about all the evolutions of Dora's complex affective life; her love for K, whom she rejects, the attraction that she feels for his wife, the reproduction in those relationships of her infantile affection for her father – her Oedipus complex-, as well as the first relationships she had with a governess. I will omit her second dream, and the scenes that marked the end of Dora's treatment.

Freud knows that the temporary lapses are important for Dora. Thus, he points to her that the 15-day period before she informs him about the end of the treatment, is the same length of time used to notify domestic staff that they are fired. And that is related – among other circumstances – with a maid who was fired after 15 days in advance when the fact that she had succumbed to her father's seduction becomes public.

Without deepening in these final scenes, I will say that Freud uses them to let us know about the *transference* concept, showing that Dora projects her previous experiences on him.

Let us now leave Freud's and Dora's story, and begin to reconstruct it using some central notions of the structuralistic conception of theories.

In what follows, I will present the principal notions of the structuralistic conception. Then I will argue within this conception about the use of the empirical claim of a theory as a strategy to reconstruct a theory, instead of the most classical approach of characterizing its mathematical models. According to the empirical claim, and following its steps, I propose a reconstruction of the applications of the *Freudian theory of hysteria* that follows the most important features of structuralism, with no the formal/mathematical structures that according to the main stream of philosophy of mathematics – and logic – are supposed to be abstract entities. In concordance with this position, a whole branch of philosophy of science sees theories as a kind of abstract entities.

14.3 The Structuralist Conception

Let us remember that for the structuralist conception the best way to characterize a theory is to specify the class of its models, and a set of applications that give empirical interpretation to the models.

For those who are not familiar with the structuralist conception and with its jargon, I will say that a theory – T – is characterized by an abstract core – K-, and a set of intended applications – I – of this core:

$$\mathrm{T} = [\mathrm{K}, \mathrm{I}]$$

In turn, the abstract core K is formed by several types of models; some of them are characterized by the non-theoretical functions of the theory, others by these functions and also by the theoretical ones, and finally, by a law-form axiom that relates the objects and functions of the theory among each other, in such a way that:

$$\mathrm{K} = [\mathrm{Mpp},\ \mathrm{Mp},\ \mathrm{M},\ \mathrm{C},\ \mathrm{L}]$$

The models characterized by the non-theoretical functions are called *partial models* – Mpp-; the models that also possess theoretical functions are the *potential models* -Mp-, and finally those that also satisfy a relational axiom are the *models* -M- of the theory.

It will be noted that I added to the models of K a relationship C that exists among the different models of the theory, and a relationship L between those models and others of other theories.

It is convenient to specify that according to the structuralist conception, non-theoretical terms are those that come from another theory, and the theoretical terms are those that are characteristic of the theory that is under consideration. It is a distinction *relative* to that theory, since a non-theoretical term in a certain theory can be theoretical in another. A function is theoretical when it can only be determined in a successful application of the theory.

It is a distinction *relative* to a theory – a function is T theoretical or T non-theoretical in a theory T – and *functional*, according to the role it plays in that theory.

It differs from the traditional distinction between theoretical and observational terms since this is *epistemological* – it refers to the observability or not of objects and proprieties – and *absolute*, since something will be observable or not in any theory under consideration.

The intended applications *I* integrate an open set of factual systems proposed as applications of the theory, and as in Wittgenstein's conception, it works as an informal semantics, giving empirical content to the abstract models of K.

In the usual characterization of the structuralist conception, the systems of *I* are subsets of the *partial models* Mpp.

If we proceeded to carry out a reconstruction following the usual standards, we would first have to characterize all the *potential models* – Mp – then cut the theoretical terms so as to identify the *partial models* – Mpp – and finally we have to establish the *models* of the theory.

Afterwards, the applications are formally identified as a subset of the *partial models*.

Of course, this approach – together with other conceptions including the traditional one – views the formal/mathematical part of scientific theories as abstract entities. However, the structuralistic conception differs from other conceptions in that it incorporates a pragmatic element to the mathematical apparatus, the *intended applications* that give empirical meaning to the mathematical models.

14.4 The Empirical Claim as a Strategy of Reconstruction

My purpose is to move away from the approach of the scientific theories as abstract entities – there is no such a thing from the physicalist point of view-, or as a non-interpreted calculus – is impossible to lose the empirical meaning when the theory is built-.

To achieve my objective of reconstructing the *Freudian theory of hysteria* without resorting to abstract entities, I will lean on what perhaps is the starting point of the structuralist approach, i.e., the solution to the problem of making an empirical claim within a theory knowing that its conceptual apparatus includes theoretical terms.

Joseph Sneed holds that it cannot be done directly, i.e. it cannot be said that "a is an S" being "a" a physical system and "S" a theory that includes theoretical terms. The solution he proposes consists in saying that a description of a physical system, characterized by means of the non-theoretical functions, can be expanded *theoretically* by adding the theoretical functions, and finally that it will satisfy also the laws of the theory. This can be said, and it does not involve any theoretical terms from the beginning. I won't develop it any longer, but I will say that this is a modification of the Ramsey-eliminability of theoretical terms.[11]

This is more than a formal solution to one of the problems outlined in empirical philosophy of science by those terms; the most interesting aspect of this proposal is its pragmatic consequences, which were immediately accepted by Sneed, in the sense that it constitutes a reconstruction of how a theory is in fact used, and consequently, it is an elucidation of scientific practice. According to the structuralist approach, when scientists do research they follow successively each one of the steps stipulated by the empirical claim.

I will go further with this statement, stating that the empirical claim reconstructs both how a theory is used, and also the stages followed when a theory is created – its genesis – since the method of construction of a theory consists mainly in making sure that an empirical system described by a previous theory can be described accurately by means of the terms that are introduced by the theory that we are considering, and that it behaves as its laws predicts.

In accordance with the empirical claim, I will keep two distinctions in my reconstruction. The first one is the difference between theoretical and not-theoretical terms. The second one distinguishes among the different structures of the theory – partial, potential, and actual-. With them I will establish the structural features that Freud uses to describe Dora as a *psychoanalytic hysterical*.

I will also conserve in the theory the role of a paradigmatic example that comes from Wittgenstein, and that the structuralistic conceptions uses as informal semantic of applications for the mathematical core, so that when a scientist finds some kind of resemblance between a paradigmatic case and another one, he is authorized to do some research to include it in the theory.

14.5 Toward a Reconstruction of Applications

Having come to this point, it is necessary to state that I intend to confer a central role to the empirical claim of the structuralistic approach. I will use its distinctions, its constructive stages and its Wittgensteinian semantics, to reconstruct the *Freudian theory of hysteria* starting from its applications. That is to say, the cases of hysteria that he analyzes, namely Dora, his paradigmatic case. This perspective could well be called *physicalist*, since I do not use any mathematical models, and all its elements are located in space and time.

In so doing, I will show a paradigmatic example of structuralism without abstract entities from the ontological point of view. And I will also show a route to be followed for those who think that Platonism is a doctrine with serious

---

[11] Sneed, J. (1971), Stegmüller. W. (1973).

epistemological problems – it has no plausible theory of knowledge-; it postulates ghost entities, and it blocks questions that are genuine from the perspective of neurosciences, because this is a scientific discipline whose research program, as we know, is in fact based on setting the foundations of an organic explanation of behavior and psychological phenomena, as neurologist Freud aptly perceived over a century ago.

One of the consequences of making a reconstruction of the theory by means of its applications is that it is always clear that although some general forms are adopted for expositions reasons – *applications*, for instance-, we are always speaking of Dora, and, consequently, of any patient with characteristics similar to Dora's.

I will show also that it is not necessary – and not even desirable – to reconstruct the theory beginning with mathematical models. This procedure entails the danger – as indeed is the case – of thinking that the latter are empirically empty, and therefore in need of some kind of interpretation. Doing anything of the sort would mean relapsing unnecessarily into an ontological Platonism, hiding that in its genesis the theory never lacked empirical meaning, and that the process always proceeds from what is closer to experience, to more general forms.

Although traditionally the word "application" is used to name facts – data – that the theory has to explain introducing theoretical terms and laws, I prefer to keep this terminology to describe also *potential* and *actual* cases so as to emphasize that they *are* physical systems, and not mathematical models.

It is necessary to remark that when I do so, I modify the standard version of the empirical claim, because I intend to specify first the *partial applications* of the theory – *Ipp*-, then the *potential applications -Ip*-, and then the full, *actual applications -I-*.

In the usual empirical claim the *partial applications* after being "theorized" – i.e. enriched by theoretical terms – become *potential models*. This sequence implies an unjustified ontological and linguistic leap. It is impossible that the addition of theoretical terms might transform an empirical application into a mathematical – abstract-structure, or an interpreted language into a pure, not interpreted mathematic language. The names of individuals and non-theoretical terms empirically interpreted in the *partial applications* do not disappear with the addition of theoretical terms – and even the latter have empirical meaning in a precise application-. Therefore, the *potential applications* are also physical, because they are the *partial* – physical-*applications* plus theoretical *interpreted* terms, added by Freud – or any other theoretician – as a hypothesis to be corroborated by the axioms that make them *actual* – physical-*applications*.

In this sense, the reconstruction of the theory based on its applications is physicalist. A natural consequence of this strategy, since even the *partial applications* that do not admit the addition of theoretical terms – and remain as such – are also physical systems.

What happens then with the psychological entities of Freud's theory, and even with the knowledge that he elaborates as long as Dora's treatment advances – or our knowledge of Freud's writings or about philosophy of science-? Are they all mental, and therefore immaterial, non physical events?

My answer is that the psychological events are basically the expression of neurophysiologic phenomena, as Freud points it out, and with whom I mostly agree on these matters. Much earlier, in *Project for a Scientific Psychology*[12] he proposes a neurological approach to the psychic apparatus; in Dora's Case (p. 113) he insists that "It is the therapeutic technique alone that is purely psychological; the theory does not by any means fail to point out that neuroses have an organic basis – though it is true that it does not look for that basis in any pathological anatomical changes, and provisionally substitutes the conception of organic functions for the chemical changes which we should expect to find but which we are at present unable to apprehend". Or: "No one, probably, will be inclined to deny to the sexual function the character of an organic factor, and it is the sexual functions that I look upon as the foundation of hysteria and of the psychoneuroses in general." There is no doubt that for Freud his *theory* is neurophysiologic, and that only his *therapeutic method* is psychological.

In brief, and more formally, we will characterize the *Freudian theory of hysteria* by means of the identification of three instances in its applications: *partial*, *potential* and *actual*, such that:

$$\mathrm{TFH} = [\mathrm{Ipp},\ \mathrm{Ip},\ \mathrm{I}]$$

Further on, we shall see that it is necessary to add to these three instances some elements that act as constraints – C- in the structuralistic conception so as to connect the different patients – cases – suffering hysteria to each other. I will complete the description of the theory by specifying its relationships with other theories, the links L. I had mentioned that the *partial applications* Ipp of hysteria – defined by the non-theoretical elements – generally come from previous theories; they constitute the facts that test the predictions of the theory, and that it tries to explain.

14.5.1 Partial Applications Ipp of Hysteria

Within the *Freudian theory of hysteria*, let us consider that the non-theoretical elements come from the psychiatry of that time, that had already characterized and identified hysteria, and of the contemporary psychology, that normally uses the representation[13] concept. These are the basic theories that underlie Freud's approach to hysteria. We know that when Freud begins to study hysteria with Breuer, he had

---

[12] Freud, S. (2000) Vol. I (1886–1899). *Pre-Psycho-Analytic Publications and Unpublished Draft* 283–294. "Project for a Scientific Psychology" The original carries no title.

[13] The representation notion was used by philosophers, psychologists and epistemologists in the Vienna of Freud's times, especially by Hermann von Helmholtz, Ernst Mach, Heinrich Hertz or Ludwig Wittgenstein. The German words that Freud uses are *Bild*, literally image or design, and often translated as picture, but he uses more specifically *Vorstellungen*, an equivalent of the idea of the British empiricists, as a representation of sensations in a private sense, or *Darstellungen*, used in a more public, linguistic sense.

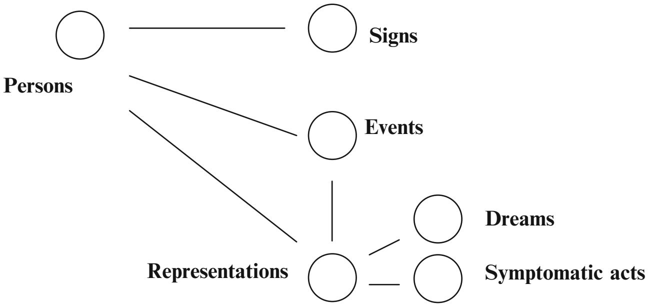
already learned with Charcot the psychiatry and the advanced neurology of his time, and particularly a key illness for the specialty, hysteria.[14] He had also concluded his stage of neurophysiologic investigations – mainly the studies on aphasia – that marked him so deeply that he never altered his claim that psychological diseases have organic bases in the central nervous system.

Let us begin then with the informal characterization of the *partial applications* – *Ipp* – of the *Freudian theory of hysteria*.

I will do it by means of a diagram which indicates the non-theoretical elements of the theory, and the relationships they establish. As a general rule for diagrams, elements are shown by means of circles and relationships by the lines joining them. The reasons to prefer this way of showing the structure of a theory are on the one hand, to make it easier to understand its main features by whoever is not too familiar with the symbols of group – and model – theory. On the other hand, by refusing to use the habitual notation of structuralism, I strongly indicate that it is not a matter of abstract models, but of empirical applications.

Of course, this exposition of the structure of the theory could be replaced by the usual set theory notation within the framework model theory, *only if we give them a physicalist foundation* such as taking them only as written signs that are understandable for a trained individual. Should we consider them in this way, it would mean, briefly, that we use a given language for pragmatic reasons, which allows us to speak of elements and relations that only can be expressed in that language.

In fact, it is sustained that the description of the theory carried out by means of diagrams can be transcribed to this language without losing anything in the process, *if it is reconstructed as applications*, and not as abstract models.



In this diagram, the objects of the theory – what it talks about, its ontology – are **Persons**, i.e. human beings. The other elements such as **Events**, **Representations**, **Dreams** and **Symtomatic acts**, since they happen to **Persons**, are not ontologically independent.

---

[14] Freud, S. (1983) "Quelques considérations pour une étude comparative des paralysies motrices organiques et hystériques", in: Anch. Neurol., 26 (77), 29–43.1888-93. English version. (2000) "Some Point for a Comparative study of organic and hysterical": "M. Charcot, he was kind enough to entrust me with the task of making a comparative study of organic and hysterical motor paralyses based on the observations of the Salpetriere."

In Dora's clinical history, they are not many **Persons**: Dora herself, her father, Mr. K, K's wife, Freud. Hardly any other actors are mentioned in the drama of Dora's hysteria.

If we deepen in other clinical histories, and in Freud's theoretical papers, usually there are interpersonal relationships with very few actors.

The *non-theoretical elements* that affect **Persons** are:

i. Somatic and psychic **Signs**, as Dora's manifestations of hysteria characterized according to the psychiatry of the time, such as the ones already mentioned.
ii. **Events,** as interpersonal events with Dora as main character and finally
iii. **Representations** of those **Events**.

The relationships between Events and Representations are schematized in the diagram by a line that unites them, as well as the relationships between Persons and Signs, Events and Representations.

It is necessary to mention that in Freud's view representations – as a mnemonic footprint of events – are basically images[15] although they integrate a complex Gestalt of visual, smell, tactile or gustatory registers.

We must add to this non-theoretical elements that where already know by doctors and psychologists, two elements that where not supposed to have any relation to hysteria at all.

They are:

iv. **Dreams** and
v. **Symptomatic Acts**.

The first one was studied by Freud in his *Interpretation of Dreams*, a text that established its importance to explore the unconscious. **Symptomatic Acts** such as tics, playing with coins in a pocket, losing objects, were also studied by Freud in their unconscious meaning before Dora's Case.[16] "I gave the name of symptomatic acts to those acts that people perform, as we say, automatically, unconsciously, without attending to them, or as if in a moment of distraction. They are acts to which people would like to deny any significance, and which, if questioned about them, they would explain as being indifferent and accidental".[17]

It is interesting to notice that in this sense Freud continues the medical tradition to reinterpret and put into a semiological[18] context elements that come from folk knowledge. I do not include them in **Signs** of hysteria because a diagnosis of hysteria

---

[15] This is the way Freud presents them in *Interpretation of Dreams*. Although in this text not only the images of the dreams are related each other in symbolic associations, but also the *words* that appears in the dreams give place to symbolic associations; in this case, the association mechanisms acts mainly considering the written *form* that the words possess, and treats them as if they were only a kind of images among others.

[16] See Psychopathology of Everyday Life, Freud (2000), Vol. VI, Chapter IX.

[17] Freud (2000) Vol. VII, p. 76.

[18] *Semiology* is an old medical discipline established by Hippocrates as the science of (observable) signs of diseases that are related to internal changes of human organs. Freud, as medical doctors do, relates these signs to internal changes of the psychical apparatus.

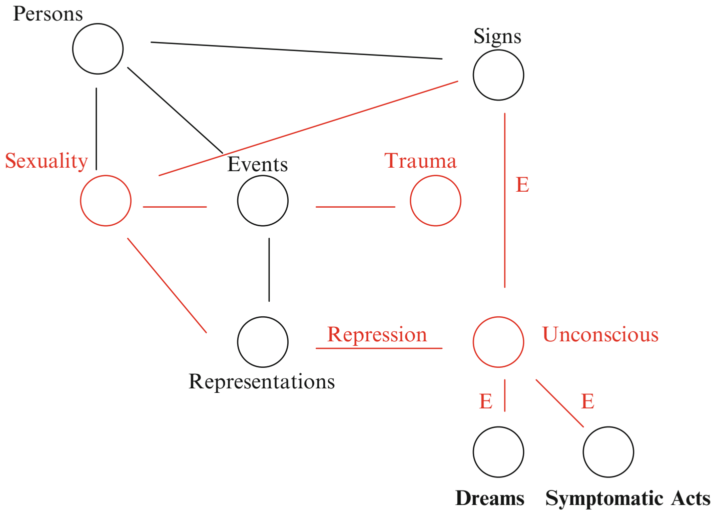
based on dreams or tics of a patient can not be done. They are non-theoretical elements that Freud uses to explore the unconscious of the patient by means of the psychoanalytical interpretations tools. They are signs of the unconscious, not of hysteria, and are related in the *partial applications of hysteria* to another non-theoretical element, namely **Representation** as it happens in normal life. We will see below that they are also an exteriorization of repressed representations of hysteria.

14.5.2 Potential Applications Ip of Hysteria

In order to obtain the *potential applications* of the theory, we add the theoretical elements to the *partial applications*.

It is necessary to remember, summarizing Freud's article about Dora that sexuality has a crucial role in the genesis of hysteria up to the point that he claims it concerns all aspects of the illness.

In a diagram:



The diagram reproduces those elements of the *partial applications* – *Ipp*-, adding to them the elements introduced by the theory. In order to help their understanding, these new elements are in red.

They are:

i. **Sexuality**. In Freud's view, sexual aspects are involved in all the features of hysteria. Actually, they act on the three non-theoretical aspects mentioned above. It appears under two forms: hetero and homosexual.[19]

---

[19] Although Freud introduces the word *libido* for the first time in 1895 to distinguish the psychological tendency –libido- from the organic sexual component –sexuality-, in the Dora's Case he always refers to "sexuality".

ii. **Trauma**. The sexual aspects of an Event transform them into a trauma.
iii. **Repression**. The action of Sexuality on Representation of traumatic Events causes their repression, and then:
iv. They become **Unconscious** for Person.
v. A relationship **E** – exteriorization-connects **Unconscious** with **Dreams**, **Symptomatic Acts** and **Signs**.

The theoretical elements of the theory of hysteria have already been made explicit. In the diagram, there are lines connecting **Sexuality** with the three non-theoretical elements. It has been stated that it acts on *Events*, and on *Representation*, and this is shown by the lines that go from **Sexuality** to **Events** and **Representation**. It is necessary to add that they also act on *Signs* since, according to Freud, when these signs attain full development, they imitates an imagined situation of sexual life; that is to say, a scene of sexual exchange, such as pregnancy, puerperium, etc. This is shown by the line that unites **Sexuality** and **Signs**.

The unconscious representations are exteriorized by means of **E**, as **Symptomatic Acts**, **Signs**, **and Dreams**.

Freud thinks that representations, as a general mechanism of the psyche, are either remembered or forgotten without causing any pathological symptom. It is only when they are connected with a traumatic sexual event – and then are repressed – that they are exteriorized by oblique roads – dreams, symbolic acts – and they are not forgotten.

14.5.3 Actual Applications I of Hysteria

Besides the theoretical and non-theoretical elements, application *I* possesses, at least one statement – axiom – that fulfills the function of relating them to each other and that is functionally equivalent to the laws of a theory. Of course they stated in a nominalistic version of a law that refers to the structure of an exemplar – a case of hysteria – and not to the set of all the exemplars – a platonic entity not admitted by our ontology-, and the possibility of using it to characterize another exemplar.

In the case of the *Freudian theory of hysteria*, those general axioms are the following:

Axiom 1:

> For any patient with signs of hysteria, there are events that sexuality -homo or heterosexual-turns traumatic; their repressed representations become unconscious, and they are exteriorized as hysterical symptoms, dreams and symptomatic acts.

We may add a methodological principle that expresses the possibility of extending the theory beyond the paradigmatic cases:

> Any other case of the *Freudian theory of hysteria* will be similar to Dora, its paradigmatic case, and it will present a similar structure.[20]

It is almost unnecessary to remark that the relationship of similarity or resemblance refers to the structures that are defined by the Freudian theory – not to Dora's physical or social peculiarities-.

These similarities should be contemplated in each application, comparing it with Dora or with other cases of hysteria, so that there is no *general*, i.e. *universal* notion of similarity.

Finally, we find in Freud' theory another very general rule that relates – associates – current representations – **Ra**-to representations of the past, and even of childhood – **Rp** - which are *privileged* unconscious representations in Freud's view.

Let us express it as a diagram:

$$\begin{tikzcd}\mathbf{R\,a} \arrow[r] & \mathbf{R\,p}\end{tikzcd}$$

I will make an additional reflection. It is implicit in the characterization of the *actual applications* that there is no inconvenience in thinking of *theoretical* cases of a theory, precisely the *actual applications I*. After all, *partial applications* are *actual applications* of a previous theory in which they possess all their theoretical functions, and in the structuralistic conception there is no doubt of its – relative-empirical character. Nothing different is postulated for the theory under consideration. Of course, this implies some kind of *internal realism*.

14.5.4 Constraints

The structuralistic conception of theories introduces some relationships among the different models of the theory that are termed *constraints*. They are proposed to unite models in a specialization and models of different specializations derived from a more general "theory element", for instance, cases of hysteria and cases of paranoia, being both of them specializations of a general theory of psychoanalysis.

In this article, I will only mention a very general relationship that unites the cases of hysteria, and allows us to apply its structures to new cases. It is also a condition of possibility of the whole psychological knowledge. It could be expressed in the following way:

"Human beings possess similar psychic apparatus, and react in an analogous way in similar situations."

Without this basic tenet there would be no psychological theories; only possible inferences about each specific social agent's performance.

---

[20] This principle is implicit in Kuhn's and structuralistic use of paradigmatic cases –following Wittgenstein's resemblance nominalism- and allows us to apply Dora's structures to other cases. We make it explicit, in a similar way as Hume does when he introduces the inductive principle.

14.5.5 Intertheoretical Links

Without intending to do a formal analysis of the theories that are more closely related to the psychoanalytical theory of hysteria, let us point out that it is related to neurophysiology – organic base of hysteria-, with physiology in general and specially with the organic bases of sexuality, with psychology and with psychiatry.

14.5.6 The Pragmatic Empirical Claim

In the pragmatic empirical claim, the steps described by the reconstruction of the theory are a kind of instructions that an epistemic subject must follow so as to use the theory as a tool for exploring the factual world.

"When Freud or another epistemic subject perceives that a person presents similar signs to those of Dora, it makes sense that he/she should investigate whether that person experienced and then forgot traumatic sexual events that are similar to Dora's. If that happens, then that person has *psychoanalytical hysteria*".

If all these conditions are fulfill a new case is added to the (open) group of patients already known.

Even when we speak of epistemic subjects, this is not a relativistic claim, since they are part of an epistemic community that trained them in shared standards, and their discoveries are tested by the other scientists. The knowledge they possess became objective – intersubjective – in the circulation among the members of the community.

14.6 Beyond Theory, Techniques

I leave out of the structure of the *Freudian theory of hysteria* two key elements of the whole theoretical frame of psychoanalysis.

The first one is the psychoanalytical *interpretation* of dreams, symptomatic acts or symptoms, because it is a tool of exploration of the unconscious – therefore a method – with the same epistemic status as other technical devices, as dissection for anatomy, the use of the telescope for astronomy, or the scales. I will add that this method possesses its own system of validation of the interpretation operated by the psychoanalyst, when the patient agrees with it, or when symptomatic acts and of symptoms disappear.

Let us remember that the roads that go from the repressed representations to their exteriorizations under the guise of dreams, symptoms or symptomatic acts follow the logic of symbolic thought, associating images to each other. Much later it will be said that the mechanism follows the rules of metaphor and metonymy. Freud points to another mechanism that does not belong to these rules, since sometimes the exteriorization appears as the opposite of what is repressed. When Dora expressed that her father is a man of (many) resources, she thinks that his resources are actually scarce, in allusion to his impotence.

Repressed representation is a Gestalt with multiple accessory facets associated to the main event that became exteriorized during the interpretation process. Such is the case with the smoke of Dora's dream that remits to the smell of tobacco in K's breath when he tried to kiss her.

The second aspect that we put outside of the theory is the *transference notion*. Freud introduces it in the end of his article – the Poscript – as a situation that occurs in therapeutic conditions, when patients unconsciously reproduce an entire series of previous events when they interact with the therapist. Freud states that it is central to interpret them so as to "avoid the delay in the cure", and thinks that because he was not aware of this notion Dora's treatment ended so suddenly.

From that moment on, it is obvious that psychoanalysis has an artificial experimental situation – the therapeutic context itself – in which to test directly its hypotheses about the past of the patient and the repressed representations analyzing the material produced in his/her relation with the psychoanalyst.

From the theoretical point of view, transference can be considered summarized by the additional axiom that relates representations of the past with events that occurs in the therapeutic relationship.

The reasons why we separate these two instances, interpretation and transference, from the theoretical body of hysteria are clear now. They are techniques and methods of exploration of the unconscious, and as such, they are not part of the theory.

On the other hand, it is the strategy followed by Freud when he refers to them in the Postscript. There he says that he also omitted all reference to the technique by means of which he extracts the content of unconscious ideas that are integrated in the total mass of spontaneous associations of sick persons.

14.7 Illness and Normality

As in the notion of illness in Claude Bernard's physiological theory, according to Freud normality and pathology are not radically different; rather, they share the same mechanisms.

According to Claude Bernard, illness occurs when there is a deviation of the normal parameters, as result of a hyperfunction or a hypofunction of an organ. It is interesting to note that the physiological method of exploration of the normal function consists mainly on extirpating the organ responsible for the function that is explored, and then observe the alterations that it originates. Usually they coincide with a defined illness; that is to say, this method causes an artificial production of an illness, in fact the one that is produced by the absence of the organ, and of course, also by its hypofunction. In the paradigmatic case of the studies on the function of the pancreas, its extirpation produces diabetes. This hypothesis has to be corroborated, and the method is completed in two ways. One of them consists in restoring the organ – perhaps as an extract – and observing if the illness is cured. The other way, is to cause its hyperfunction. In this case, that should cause the opposite dysfunctions produced by the extirpation. Again, in the case of the pancreas, the injection of insulin replaces the function of the pancreas, normalizing the levels of sugar, and the excess of insulin produces hypoglycemia.

Similarly, in Freud's view the mechanisms of hysteria and of normality do not differ, and as with the physiology of the pancreas, he can investigate the normal functions of the psyche by studying its illnesses. When Freud speaks of hysteria, he also informs us about the normal mechanisms of the psychic apparatus, and of its deviation in this illness.

It is no coincidence, then, that the correct interpretation of the material that surfaces in analysis and its corroboration by the patient, lead to the disappearance of the hysterical symptoms and of the symptomatic acts, and therefore to the patient's cure.

It will be remembered that when, following its interpretation the repressed representation becomes conscious, the patient is cured. When there is no more repression the normal psychophysiology is restored, and it causes no more symptoms.

14.8 Theoretical Background of Hysteria

Of course, this it is not Freud's first article, nor his first theoretical reflections. Even when he does not mention them explicitly – or too seldom – his assertions should be understood within a theoretical background that embraces the whole psychoanalytical theory, as it was formulated at the time. I refer mainly to the postulation of a psychic structure, the notion of libido or the Oedipus complex.[21]

Although all of those notions are assumed knowledge whose reconstruction is not the purpose of this article, I will characterize them briefly within the conceptual frame introduced up to this point.

In Freud's article there are not many elements of a psychic structure. We already saw that the notion of the unconscious appears as representations that are repressed, and that Freud obviates the notion of consciousness, simply mentioning that they are remembered otherwise.

About the libido notion that he introduced years before, we can see it as a function of sexuality. It has no significant role in Freud's article.

The Oedipus complex is mentioned very briefly when Freud speaks about Dora's relationships with her father. Without attempting its reconstruction, I will suggest that it can be understood by referring to a similar interpretive structure as the one sketched above, stating that the elements of the Oedipus complex are specifications of those of the theory of hysteria, namely people, facts, representations, repression, etc.

---

[21] The psychoanalytical theory experienced several developments and changes in the course of time. I will refer exclusively to those that happened before the article on Dora's Case.

14.9 Psychoanalytical Practice

I had already mentioned that the empirical claim of the theory expresses the possibility of extending the paradigmatic case to other cases, and at the same time it points out the steps that Freud follows when he invents-discovers the theory, and also when it is applied. That is to say, it establishes the steps of the genesis and the use of the theory.

I may not have stated explicitly enough that when I mention the characteristics of the use of the theory, I am indicating at the same time that the psychoanalytical diagnosis follows those same stages. And lastly, when this diagnosis transforms a hysterical patient into a psychoanalytical one with clinical investigation covering the phases specified by the empirical claim, the successive corroborations that the analyst receives from the patient cures him. That happens because they are of such a nature that the repressed material goes from unconscious to consciousness.

Genesis of the theory, diagnosis and psychoanalytical cure coincide in their structure with the pragmatic empirical claim that results from reconstructing the theory as applications, following the text of the paradigmatic case, Dora.

14.10 Addendum

We started our journey with a brief report of Dora's Case, and now finish it with a physicalist reconstruction of the psychoanalytical theory of hysteria.

In the curse of this analysis, we found out that Freud follows a sequence of steps when he turns Dora on his paradigmatic case of *psychoanalytical* hysteria, and notably, those steps coincide with the stages required to make an empirical claim of the theory, i.e. its use.

Although the empirical claim allows us to reconstruct the way scientists use their theory when they explore its field of application, in our analyses we notice that it is a central tool to understand also the stages of constitution of the theory, its genesis.

This is crucial to establish the impossibility of postulating that the theory can be appropriately characterized by means of abstract models – without interpretation – since it begins with empirical applications, and in the progressive construction of the theory they never lose their factual meaning. The structural core of the theory can not consist of empty, mathematical models.

On the contrary, it persuades us that if we want to be fair to what a theory is, we should stick to the reconstruction of those empirical structures that at the beginning only have nontheoretical elements, afterwards, their theoretical elements, and finally, satisfy certain law-like axioms that relate all the elements. Briefly, a reconstruction of applications of the theory that first are *partial*, then *potential*, and finally, *actual* applications, always in the level of the characterization of physical systems, and without resorting to mathematical models.

We could therefore characterize the *Freudian theory of hysteria* by means of its applications, its constraints, and its inter-theoretical relationships, such that:

$$\text{T P of H} = [\text{Ipp, Ip, I, C, L}]\ \text{Dora being I}_0$$

Such that:

i. Ipp are *partial applications*
ii. Ip are *potential applications*
iii. I are a*ctual applications*
iv. I_0 is a paradigmatic application in the initial time 0 of the theory: Dora
v. C are constraints
vi. L are inter-theoretical relationships

We need nothing else to characterize the theory.

It is unnecessary to appeal to abstract, non-interpreted structures.

A factual theory does not begin or involve the presence of those mathematical structures – models – that were thought indispensable, as the core of the condition of possibility of experience.

We also show that the structural knowledge of the paradigmatic applications is actually the condition of possibility, the one that facilitates its expansion to other cases, when an epistemic subject perceives structural resemblances among them.

Although we employ a more general language using letters such as *I*, *Ipp*, words as *applications*, *links*, etc. they do no point to an abstract entity, we are always describing Dora, the paradigmatic application, and by doing so, any other case of hysteria, including fictional ones. Perhaps this is the last stage of the process of invention of theoretical terms, simplification and generalization that begins with the notes that Freud writes during the treatment, continues with his article and finishes with our reconstruction. This process goes from the details of a description, to the most general forms of philosophy of science.

Upon doing so, we come to a kind of language and of structure that allows us to describe any possible case of hysteria, including fictional ones. This does not differ from what happens in physics, or other disciplines, when students are trained by means of fictional exercises, and do not need to test the results in laboratories.

When we speak of fictional cases, we do not deny physicalism. They exist as physical events in the form of thought-language[22] of the epistemic individual that proposes them, or as statements spoken or written in a paper.

Not only fictional cases challenge nominalism. The presence of *potential applications* in the reconstruction outlines the problem of its ontological status, since they

---

[22] I use thought-language in the same sense than Neurath (1983, p. 67) when he states: "We speak not of "thinking" but straight away of "speech-thinking", that is, of *statements as physical events*." This is a general physicalist solution to ontological problems of language –and of logic and mathematics-. This strategy denies the existence of abstract entities such as types, ideas, etc., proposed by Platonists. Of course, some statements refer to physical systems and others do not refer at all.

might be interpreted as a possibility that refer to non-spatiotemporal entities. From the physicalist point of view, there is no such a thing as non actual physical systems. The presence of theoretical elements in a patient – or in any physical system – is a *hypothesis* stated by the psychoanalyst, and justified if the patient follows the laws of the theory. It is no only the case that this hypothetical statement is as physical as the fictional ones; it differs from them because it refers *to a physical system* and not to a mathematical entity and therefore keeps in the process this ontological condition.

It may be noticed, just as I said at the beginning, that central elements of the structuralistic approach are kept in the reconstruction, expressed in physicalist terms –and in physicalist ontology – fulfilling Wittgenstein's nominalistic strategy that is in the basis of the semantics of structuralism, and in the application of a theory. We also saw that this nominalistic use of the theory coincides with the diagnosis, the treatment, and the recovery of the patients, since turning a hysterical patient into a psychoanalytic one, implies his full diagnosis as such. Its corroboration by means of becoming conscious of the repressed representations restores the normal physiology of the system, and therefore the recovery of health.

Although this view diverges from the standard presentation of the structuralistic conception, it does not constitute a radical heterodoxy.

From time to time, we read manifestations that point in the same direction, although they are not fully developed.

Let us see how C.Ulises Moulines (1998, p. 154) states it:

> And if authors like Goodman, Field and others are right, then in factual theories we never need those terms (the mathematical). We could do without them when dealing with knowledge of empirical reality (that is to say with "genuine" knowledge) and we could spare ourselves the metaphysical headaches they represent for us.[23]

He stated before (p. 148):

> It is not absurd to imagine that the essential aspects of SMM (Structuralistic Meta-theory Methodology) can be reproduced in the general frame of a (strong enough) nominalistic or intuitionist system.

We coincide totally with these words. We believe, indeed, that "we can do without them (the mathematical structures)", and it is not absurd to imagine a reconstruction of a theory in the general frame of a nominalistic system.

Also we coincide with the comment he makes further down, concerning the ontological nature of mathematics (cit. p. 154):

> The debate is still open, and this is not the place to risk anything substantial about it.

Our proposal constitutes a reply to those concerns, taking a decidedly physicalist stance, relegated by the Platonic positions that predominate unnecessarily in philosophy of mathematics and logic, and in consequence, sometimes in those philosophies of science that employ formal tools to elucidate the structure of science.

---

[23] English version of CL.

This is our homage to Otto Neurath. We propose an updated physicalist view of contemporary philosophy of science, using physicalistic and nominalistic tools to reconstruct a factual theory. In so doing we keep the distinctions that have been forged since Neurath's time and that cannot be neglected by philosophers of science who are interested in the structure of scientific knowledge.

References

Appignanesi, L., and J. Forrester. 2000. *Freud's women*. London: Penguin.

Balzer, W., and P. Marcou. 1989. A reconstruction of Sigmund Freud early theory of the unconscious. In *Psychological theories from a structuralist point of view*, ed. H. Westmeyer, 13–33. Berlin/Heidelberg/NewYork: Springer.

Balzer, W., C.U. Moulines, and J. Sneed. 1987. *An architectonic for science*. Dordrecht: Reidel.

Brambrough, R. 1966. Universals and family resemblance. In *Wittgenstein*, ed. G. Pitcher, 186–205. New York: Anchor Books.

Carnap, Rudolf. 1956. Empiricism, semantics and ontology. In *Meaning and necessity*, ed. Rudolf Carnap, 205–221. Chicago: The University of Chicago Press, Enlarged Edition; 1958. The methodological character of theoretical concepts. In *Foundations of science and the concepts of psychology and psychoanalysis*. Minnesota Studies in the Philosophy of Science, vol. I, 38–77.

Diez, J.L., and C.U. Moulines. 1997. *Fundamentos de la filosofía de la ciencia*. Barcelona: Ariel Filosofía.

Field, Hartry. 1980. *Science without numbers: A defence of nominalism*. Princeton/Oxford: Princeton University Press/Blackwell.

Frege, Gottlob .1892. Uber Sinn und Bedoitung. In *Zeitschrift für Philosophie und philophische Kritik*, vol. 100, 25–50, ed. P. Geach and M. Black, *Translations form the Philosophical Writings of Gottlob Frege*. Oxford: Basil Blackwell; 1918. Der Gedanke. In *Beiträge zur Philosophie des Deustchen Idealism*, I; Translation of A.M. and M. Quinton. 1956. The thought, a logical inquire. *Mind* 65(259): 289–311.

Freud, Sigmund. 2000. *The Standard Edition of the complete Psychological Works of Sigmund Freud*. Translated from the German under the General Editorship of James Strachey in Collaboration of Anna Freud (24 Volume set). The Hogart Press and the Institute of Psychoanalysis, London 1953–1974.

Gruenbaum, Adolf. 1984. *The foundations of psychoanalysis*. Berkeley: University of California Press.

Goodman, Nelson. 1972. *Problems and projects*. Indianapolis/New York: The Bobbs Merril Company, Inc.

Goodman, N., and Quine, W.V. 1947. Towards a constructive nominalism. en: *Journal of Symbolic Logic*, 12: 105–122.

Lorenzano, César. 2001. Scientific theories, ontology and languaje. In *Formal theories and empirical theories*, ed. J.M. Sagüillo, J.L. Falgueras, and C. Martinez, 623–637. Publications of University of Santiago de Compostela; 2002. La estructura pragmática de la ciencia. In *Communication to the third meeting of structuralistic metatheory*, Granada.

Moulines, C. Ulises. 1982. *Exploraciones metacientíficas*. Madrid: Alianza. 1996. Structuralism: The basic ideas. In *Structuralistic theory of science*, ed. W. Balzer and C. U. Moulines. Berlin: Walter de Gruyter; 1998. Esbozo de ontoepistemosemántica. In *Teoría*, Vol. 13, Número 31, January 1998, San Sebastián, Spain, 141–159.

Neurath, Otto. 1931a. Physicalism: The philosophy of the Viennese circle. *The Monist* 41: 618–623.

Neurath, Otto. 1931b. Physicalism. *Scientia* 50: 297–303.

Neurath, Otto. 1983. *Philosophical Papers 1913–1946*. Trans. and ed. Robert Cohen and Marie Neurath, *Vienna Circle Collection*, vol. 16. Dordrecht/Boston: D. Reidel.

Peirce, Charles S. 1966. Letters to Lady Welby. 31 Jan 1909, 406. In *Charles S. Peirce: Selected Writings (Values in a Universe of Chance)*, ed. Philip P. Wiener. Dover: New York.

Rodríguez-Pereyra, Gonzalo. 2000. *Resemblance nominalism. A Solution to the Problem of Universals*. Oxford University Press: Oxford.

Sneed, Joseph. 1971. *The logical structure of mathematical physics*. Dordrecht: Reidel.

Stegmüller, Wolfgang. 1973. *Theorienstructuren und Theoriendynamic*. Heidelberg: Springer; 1979. *The structuralistic view of theories*. Berlin: Springer.

Suppes, Patrick. 1960. A comparison of the meaning and use of models in mathematics and the empirical sciences. P. Suppes. 1969. *Studies in the methodology and foundations of science. Selected Papers from 1951 to 1969*, 1024. Dordrecht: Reidel.

Wittgenstein, Ludwig. 1958. *Philosophical investigations*. Oxford: Basil Blackwell; 1956. *Remarks on the foundations of mathematics*. Oxford: Basil Blackwell.

Chapter 15

The Cultural Sciences and Their Basis in Life. On Ernst Cassirer's Theory of the Cultural Sciences

Christian Möckel

Abstract The cultural philosopher Ernst Cassirer campaigns for the idea of a unity between natural and cultural science. He polemizes, however, against the idea of the 'unified science' as proclaimed by the 'Vienna circle' which favoured physics and the concrete sense perception as the only forms of objectification. The personal-emotional experience of expressive perception which constitutes forms of culture and is therefore the starting point for any cultural science engagement is considered by Cassirer to be an independent and alternative form of objectification of the mind. In his opinion, expressive perception genetically precedes the perception of sense qualities. Both types of science are connected by the relations they draw between the singular and the general. They can be differentiated by their targets – respectively objective laws and subjective shapes or forms. By referring to the emotional expressive perception Cassirer insolubly connects the 'primary phenomenon of life' and its appropriate characteristic of the mental life with concept formation within the cultural sciences. His theory of "basis phenomena" as the final modes of the constitution of reality within perception provides a framework for concept formation that is centered around the "basis in life" and the "life feeling". Forms of life, physiognomic imprinting of sense and epochal character of style are associated in the concepts of the cultural sciences with a specific demand for generality and objectivity which simultaneously differentiates them from the concepts of natural science whilst also equating them.

Keywords Basis phenomena • Concept formation • Cultural sciences • Expressive perception • Objectification • Phenomenon of life • Symbolic form

---

C. Möckel (✉)
Institut für Philosophie, Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany
e-mail: MoeckelC@philosophie.hu-berlin.de

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology, and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_15,

15.1 Introduction and Tasks

In this paper I would like to demonstrate how the cultural philosopher Ernst Cassirer (1874–1945) in his writings during his time in Gothenburg (1935–1941) bases his considerations within the philosophy of science on the specific concept formation in the culture sciences upon the primary phenomenon of life (Urphänomen des Lebens). This is currently well-known material within research on Cassirer, but outside this field this still little considered fact regarding a back reference of scientific concept formation to the so called 'Urphänomen' of life offers an original approach to philosophy of science. This is done by basing the idea of an independent objectivity in the cultural sciences on the most elementary form of expression of human life: perception. The theory of the cultural sciences not only gains a natural basis by this anchorage in the primary phenomenon of life that works against one-sided idealistic assumptions, but also acquires – by the physiognomic phenomenon of expression that is immanent in every living thing – the specific conceptual means to understand culture as a symbolically formed and arranged diverse universe of sense. Cassirer hereby contributes to the incessant philosophical problem of thinking of reason and life not merely as bare unmediated contrasts but as *two* inseparable aspects or directions of *one* entity, the human faculty of expression and objectification.

This is significant as Cassirer, during his earlier time in Berlin, made use of the scientific approach of modern mathematics as developed by Hermann Cohen and Paul Natorp to account for the history of the epistemological task within philosophy and science of the modern era (Cassirer 1999 [ECW 2]). The famous 'turn' within his Philosophy of Symbolic Forms – carried out in the period between the two world wars – from the critics of reason to the critics of culture draws up a philosophy of culture founded upon the categories of form, symbol and expression. With this transition from epistemology and theory of science to a symbol theoretical philosophy of culture the former Marburg School Kantian Cassirer discovered a further basic category of modern philosophy. Despite the long prehistory (Möckel 2005, 25ff.) it is this movement towards the concept of life – as opposed to the usually conceived opposition of mind and life – as a primary phenomenon by the revaluation of the expressive perception (Ausdruckswahrnehmung) towards a key or basic concept of his philosophy in the 1920s that marks a new qualitative step (ibid. 141ff.). In the 1930s Cassirer subsequently developed the metaphysically inclined theory of the basis phenomena (Basisphänomene) of any kind of perception (Cassirer 1995 [ECN 1], 113–198) that built the reference point for his later works of a theory of the cultural sciences whilst in Gothenburg and New Haven.

Following Kants concept of unity (Einheitsbegriff), Cassirer became interested in the methodical and object-related relation between natural and cultural science and extended this questioning towards mathematics and historiography. In 1941 he concluded that the inner crisis of philosophy and science over the past hundred years since the death of Goethe and Hegel was at least partly the result of the unsolved relation between natural and cultural science (Cassirer 1989, 34). Despite his will to overcome the gap between natural science, cultural science and science of history that had widened during the nineteenth century, he did not support the positivist methodical monism of Julius Kraft (Cassirer 2004 [ECN 5], 33). He polemized against the Unified Science (Einheitswissenschaft) as it was portrayed by the representatives of the 'Vienna Circle', especially Bertrand Russell and Rudolf Carnap, that excluded logic and mathematics and revaluated physics and sense perception as the only source of knowledge (ibid. 73ff., 176f.). He contradicts the claim that there are no *non*-physical facts of the matter (ibid. 80) by referring to the genetically preceding expressive perception, and he emphasizes the methodic differences between the intended concepts of law and form within the natural and cultural sciences – both nevertheless unified in their aim to relate the singular to the general (ibid. 132f.). These two scientific types thus share a general function of objectification of the experiences, though each having its own specific character.

15.2 Expressive Perception as a Basic Layer of the Cultural Sense

Following Cassirer the different scientific types utilize different ways to objectivize and gain knowledge about the world. The decisive methodical step is fulfilled by questioning if those ways can be reduced to some last and fundamental sources and modes of experiencing reality. By referring to Goethe Cassirer does indeed show such last modes of mediation and unlocking of reality within the most elementary way of experiencing reality: perception. Modes that constitute reality by cognition and mediation by being directed towards an emotional-personal 'You', a sensory-objective 'It' or subjective-experiencing 'I'. These modes of perception on the one hand do work unmediated, and that is why he calls them primary phenomenons or basis phenomena. On the other hand, they are only accessible to us through the reflections of the mind and thus in a mediated and incomplete manner.

From the stances of the different sciences, these primary phenomena of perception are overlooked, and it is only by a methodological scope that they become subject to the scientist (Cassirer 1995 [ECN 1], 139). Following Natorp (Natorp 1912, 189ff.) Cassirer explains that the basis phenomena of experiencing reality become apparent if we turn back from the objective constructions to their subjective sources and origins (wenn "wir von den 'objektiven' Gebilden zurückfragen nach ihren 'subjektiven' Quellen und 'Ursprüngen'" [Cassirer 1995 (ECN 1), 145]). He emphasizes that those ways of mediation cannot be studied in isolation but only as a whole and in correlation with one another.

In his philosophy of science, Cassirer gives reason for his decision to turn to an analysis of perception as a basic layer of all phenomena of consciousness by pointing out that an analysis of the concept formations barely contributes to an understanding of the objects of the cultural sciences and their cultural character. The task as he sees it should be, on the contrary, to track down the concepts of the cultural sciences - as well as the object formations within the natural sciences – to their fundamental source (Cassirer 1989, 56). It was already in 1925, in his work on *Mythical Thought* (*Das mythische Denken*) (Cassirer 2002 [ECW 12]), that he discovered the emotional expressive perception as the most original intelligible-sensory performance upon which the other theoretical performances are based. By reference to the three basis phenomena he is now able to describe and understand these settings and their directions phenomenologically.

The description of this phenomenal inventory of the elementary perception reveals their intentionality in two directions, both towards 'You' and towards 'It' (Cassirer 1989, 39). Depending upon the direction of intentionality, perception gains a special colouring and hue, i.e. the experience of reality gains the meaning of personal expressive perception or of thing perception (ibid. 40). Embedded in both directions of perception there is – as a third direction – the possibility of a relationship towards the 'I' as its bearer and starting point (Cassirer 1995 [ECN 1], 122). The original expressive perception, in which both intentional directions are still unified and which is characteristic of mythical consciousness particularly, proves for Cassirer a primary phenomenon of the living one, an ultimate basic and original layer of all mental energies and thus also a last source of knowledge. It is within this expressive perception that Cassirer situates the rootage of the different scientific term types in the primary phenomenon of life. At the same time he recognizes that both philosophy – in the shape of logical positivism – and the natural sciences in accordance with their own logic seek to clog the source, from which the myth constantly feeds, by denying expressive perception a right of its own (die "Quelle zu verstopfen suchen, aus der der Mythos sich ständig nährt, indem sie der Ausdruckswahrnehmung jegliches Eigenrecht bestreiten" [Cassirer 1989, 40]).

Cassirer's thesis is that expressive qualities are then replaced by sense qualities and these subsequently by pure quantitative terms (ibid. 40f.). In this way the theoretical view of nature has displaced anything 'personal' and switched it off (ibid. 48). Cassirer does not want to reverse this tendency to turn away from their basis in life which is somehow constitutive, however, the understanding of reality resulting from it does provide a reason for concern. The pure goal of material objectivity in the natural sciences nevertheless proves particularly reality-narrowing if we have to deal with the built world of humans, that must be understood both as natural and through cultural 'It' – objectivity. The true works of culture unite both kinds of objectivity in themselves; it is at this level of material objectivity that the cultural senses are to be recognized and explained.

Cultural life, to which among other things language, art, religion and state belong (ibid. 50), reveals itself in the works of culture. By means these works, the material-physical dimension alongside the psychological dimension of the creating individuals and the social (historical) dimension of the lives of these individuals can be distinguished. Although physical, psychological and sociological terms are able to recognize each of these elements, they miss the actual cultural sense of the works, i.e. their religious, artistic or scientific sense. These we exclusively understand with the help of the independent concepts used by cultural science as a virtual fourth dimension. These concepts are only accessible to us if we consider the subjective, live and personal hue with which the expressive perception impresses upon on all its contents. This is why it plays a crucial role in the organization of cultural objects.

This leads us, according to Cassirer, to the understanding that works of culture receive their idealistic moments which constitute their specific sense and their form as a religious, aesthetic or scientific work through a personal imprint. This imprint is executed upon the sensory-material embodiment of a presented content. There are, therefore, three dimensions that determine a work of culture: the physical existence, the thing represented and the personal-expressed (ibid. 42f.). The expressive perception as the primary phenomenon of life proves to be an appearance which is not amorphous, but has direction, form and structure.

Historiography – which was for Cassirer one of the preconditions for cultural science – incorporates the all animating expressive character and utilizes it. It is with the assistance of the expressive language (physiognomy) that the characteristic of the cultural, including the national "life of an epoch", is unlocked (ibid. 82). With this necessary theoretical reflection, however, all direct, intuitive security and certainty of the physiognomic is lost (ibid. 45). Cassirer, however, insists on not banishing methodically "the whole richness of pure expressive experiences" from either historiography or the cultural sciences (Cassirer 1999 [ECN 2], 14, 16f.).

The pure expressive experience therefore belongs to "the constituents of just this 'world'" (ibid. 138f.). It has finally turned out to be an ineradicable primary phenomenon on an "imperturbable basis". All forms of culture and cultural senses such as language prove from this point of view to be lively forms that have their basis in a life-phenomenon. The expressive perception reveals its "legal basis" (Rechtsgrund) particularly within the structure of the various worlds of cultural senses. It provides here a "distinctive and independent source of light" which we should not miss if we want to make "the structures of the cultural world transparent" to ourselves. Expressive perception proves that it belongs to "the basic means of objectification" and thus possesses "a genuine objectivity" (ibid. 151). It is therefore to be treated as "the natural starting point for all research within the cultural sciences". This provides a basis on which the concept of a cultural science expresses a characteristic that binds it to a "'physiognomic' realization" (ibid. 165, 168).

15.3 Concept Formation in the Cultural Sciences

In order to gain a more detailed insight into the characteristics of the concept formation within the cultural sciences and its anchorage to the personal 'You'-basis phenomenon, i.e., in the original-intuitive experience of the other, it is always necessary to keep in mind the constitutive functions of the other two primary or basis phenomena involved with the perception of reality. In Cassirer's theory all three primary phenomena take part (in different and specific ways) in the constitution of the individual types of science or the directions of objectification. The demand to give "a general, systematic overview of the different methods" applied "in the single sciences" that "constitute their 'concept of reality'" (Cassirer 1999 [ECN 2], 186) was supposed to be fulfilled in a work that did not appear during his lifetime.

The "basic directions of the cognition of reality" that can be attributed to the three basis phenomena and which concern themselves with three basic types of science have their effects exemplarily in abstract mathematics, in the accurate knowledge of nature (theoretical physics and descriptive biology), and in the culture sciences including historiography. Cassirer also stresses the independent ideographic motive of knowledge addressed within historiography. Within all three directions the theoretical objectification process, i.e., "the world of the concept and of the scientific knowledge", builds not only on empirical intuition, but also on the inseparability of expressive and sense perception (ibid. 31). In their resulting states all forms or directions of reality cognition show a certain type of intended concreteness. These intentions fulfill themselves in different ways with regard to the three basic directions of objectification. It is only "the form of the object relation as general form" that remains unchanged (ibid. 29).

According to Cassirer, the primary phenomenon of the immediate, flowing subjective life ('I') can be promoted to that of pure producing and designing within mathematical thinking. The specific manner of knowledge within the mathematical synthesis, the mathematical construction of certain subject areas, fulfills a pure possibility of thinking. It does not claim a counterpart in the mental or sense-empirical reality. Modern mathematics trains a "close intellectual cosmos", one of particular objectivity in which the natural numbers do not have any priority (ibid. 67). With its specific epistemological motive, mathematical thinking acquires general terms by constructive deduction, it does not however acquire species concepts by induction.

The two other forms of reality cognition share a common goal, namely, to extract "certain organizations which return in homogeneous ways" from "the flow of the happening" which is experienced by an 'I'. The 'It' – direction of the synthesis which leads to the objectifying sense perception organizes the empirical and finally theoretical order of the world as a work of humans. It leads to objective laws (ibid. 92), since this motive for knowledge aims at scientifically intended constants and invariants in "the flow of the happening". It is upon the empirical term shaped by this direction that the accurate term of mathematical natural science is based. It is with mathematical knowledge that the modern natural science transcends the empirical-descriptive world (ebd, 109); their specific invariants constituting a new concept of reality directed at the perceivable world (ibid. 113).

The 'You'-direction of the synthesis, however, which points towards the all-animating expressive perception with its personal configurations is directed towards physiognomic forms (ibid. 83f.). This third direction of the reality cognition, according to Cassirer, should by no means only be treated as a remainder of early mythical thinking that is supposed to be overcome. Rather it forms its own method of objectification within human performance, which alone shapes the forms of cultural sense. It has therefore to serve in the cultural sciences as a starting point and is supposed to provide an appropriate way of concept formation. The perspective of the cultural sciences is targeted as an epitome of mental 'forms' and their understanding (ibid. 174). The concepts are determined here as “form and style concepts”, which detect “structure problems”. They can therefore be neither ‘nomothetic’ nor ‘ideographic’ concepts (ibid. 58f.), as Wilhelm Windelband labeled the two possible types of concepts (Windelband 1919, 145f.). These structure concepts all belong “to the same logical ‘family’”, which differentiates them from the concept of natural law within the natural sciences. Each concept of style, structure or form expresses however a “‘world view’ of its own”, encompassing its own “basic direction of thinking and imagination”, of “looking at and seeing” (Cassirer 1999 [ECN 2], 62).

The problems presented by the basic directions in accordance with the concept formation of the three primary phenomenal mental behavior types can also be approached in terms of “the [three] circles of being”. The original circle for Cassirer is the ‘I’ – mode in its live abundance and restless motion. This ‘I’ however experiences in the external world of the other (‘You’), as a second circle, a limitation of its doing (ibid. 9f.). A third limiting “circle of being” is experienced by the ‘I’ in the actively created works (‘It’) in which its acting gives itself a form. This objectivated being originates, Cassirer states correctly, “completely from us” whilst “nevertheless not belonging to us”. As “specific-human reality”, or “reality of the culture” (ibid. 10f.) it covers also the natural sciences including their objects of knowledge. It is at this point that the world of culture differs significantly from that of nature. The cultural sciences pursue a specific goal of inquiry in this third circle, which “constitutes a third independent axis of coordinates and thus establishes a new ‘dimension’ of knowledge” (ibid. 153). They aim to recognize “the totality of forms in which human life takes place” (Cassirer 1989, 76).

As the objects of cultural science are described and interpreted historically, Cassirer pays much attention to the methods and the methodical self-understanding of the historiography of culture. Although the organic forms of nature are to be treated evolutionarily, it is the system of culture, the system of the symbolic cultural forms that constitutes for Cassirer the center of history. The historian, who provides the material for the cultural scientist – and the philosopher – enables an “insight into the life forms of the past” (Cassirer 2000 [ECW 5], 361). The science of history seeks to understand “past life” by interpreting it, thus retaining its form, as the contents cannot be renewed and revived (Cassirer 1989, 77). The historical view of the cultural forms of human life functions as a material basis for the cultural sciences that unfolds and subsequently investigates these forms. Cassirer attributes to the individual cultural forms the same subjective-objective, sensory-senseful double nature, that also suits the symbolic forms. The cultural “shape” is and remains connected to the “primary phenomenon of life” via this double nature and its rootage in emotional expressive perception (Cassirer 2000 [ECW 5], 343). Almost any cultural shape therefore contains the basic conflict between the empirically historical and the idealistically timeless.

As the cultural sciences achieve a revival of the culture of the past with the help of “the richness of the concepts of form and style”, the historical monuments which are investigated by historiography are understood by them as symbols. They let us recognize and restore “certain ways of life”. “The cultural sciences teach us to interpret symbols and to unravel the mysteries that are locked within them". Thus making "the life from which they originated visible again" (Cassirer 1989, 86). The productive process of creating cultural symbols and the reflexive process of understanding them form necessarily opposing directions of mental activity that cannot be carried out simultaneously.

Each form investigated in the cultural sciences expresses a concrete character of a cultural phenomenon and coincides therefore with the appropriate symbolic form. The individual culture sciences are therefore theories of a specific symbolic or cultural form. Through "universal outlook" the particular method of any discipline of cultural science extracts something archetypical from the richness of the phenomena. This 'ideal-typical' view (Max Weber) forms, for Cassirer, its "own and legitimate kind of concept formation in the cultural sciences" (Cassirer 1999 [ECN 2], 161f.). The possible cognition of the archetypical "basic forms of culture" presupposes a certain perspective or way of seeing. The concept of the whole, an epoch etc., expresses something characteristic that leads us back to a "'physiognomic' realization" (ibid. 168).

For the study of the works of culture, specific methods of cognition are demanded which unfold the material and the form moment at the same time. The concepts which are supposed to clear the culture objects and their idealistic "stock" for Cassirer are understood in their specifics from the way "the subsumption of the special under the general is executed". This requires another "kind of 'synopsis'" than those within the natural sciences (Cassirer 1989, 72f.). Even if the synopsis is common to both types of science, it is a different relationship between singular and the general than the one that is aimed at by causality and the laws of natural science. This specific relationship belonging to the nature of the concept of the cultural sciences Cassirer calls "the life-thread of the concept" (Lebensfaden), that may never be severed (ibid. 69f.).

This specific way of subsumption is always related to an appropriate "life feeling" (Lebensgefühl). The view of cultural science aims primarily – however not exclusively – at the subjective, since it differentiates between the aesthetic, religious and legal area of activity. Cassirer shows that every form of culture that comes into the awareness of a cultural science is deeply rooted in the subjective life feeling of the individual. He stresses the fundamental idea of the subjective life feeling as the primary layer of rational behavior. He was in this point inspired again by Goethe who had been led to his theoretical insights by this particular life feeling (Cassirer 2000 [ECW 5], 163, 166).

Finally, Cassirer's argument against cultural nihilism must be addressed. Since "the world of the forms" can only be recognized in "their functional execution" and as building movement of the 'I' (Cassirer 1999 [ECN 2], 170) the subject may not lose itself within this perspective. It would lose all objects and forms by losing the organizer of these forms. This is why in the cultural sciences it is not the finished work but the respective form-building activity of the subject that constitutes the main center of interest. With regard to the culture form which does not destroy the mental being but rather functions as a condition for it, Cassirer uses the Goethean notion of metamorphosis as a suitable term (ibid. 172f.). Cassirer sees the problem of the work that is unlocked by the 'It' – basis phenomenon in which humans objectivate themselves and become mediately recognizable by others, brought to the fore in both: Simmels theory of 'the tragedy' of culture and in Goethe's poem *Prometheus* (Cassirer 1995 [ECN 1], 125). But this is another topic which requires lengthy investigation.

References

Cassirer, Ernst. 1989. *Zur Logik der Kulturwissenschaften*. Fünf Studien [1942]. Darmstadt.

Cassirer, Ernst. 1995. Über Basisphänomene. In Zur Metaphysik der symbolischen Formen. Hg. von John Michael Krois. In *Nachgelassene Manuskripte und Texte*. Hg. von John Michael Krois und Oswald Schwemmer, Bd. 1: Hamburg. (ECN 1), 113–198.

Cassirer, Ernst.1999. Das Erkenntnisproblem in der Philosophie und Wissenschaft der neueren Zeit. Erster Band (1906). In *Gesammelte Werke*. Hamburger Ausgabe. Hg. v. Birgit Recki, Bd. 2, Hamburg (ECW 2).

Cassirer, Ernst. 1999. Ziele und Wege der Wirklichkeitserkenntnis [1937]. Hg. von Klaus Christian Köhnke und John Michael Krois. In *Nachgelassene Manuskripte und Texte*. Hg. von John Michael Krois und Oswald Schwemmer, Bd. 2: Hamburg (ECN 2).

Cassirer, Ernst. 2000. Das Erkenntnisproblem in der Philosophie und Wissenschaft der neueren Zeit. Vierter Band. In *Gesammelte Werke*. Hamburger Ausgabe. Hg. v. Birgit Recki, Bd. 5: Von Hegels Tod bis zur Gegenwart (1832-1932) [1957], Hamburg (ECW 5).

Cassirer, Ernst. 2002. Philosophie der symbolischen Formen. Zweiter Teil: Das mythische Denken [1925]. In *Gesammelte Werke*. Hamburger Ausgabe. Hg. v. Birgit Recki, Bd. 12, Hamburg (ECW 12).

Cassirer, Ernst. 2004. Kulturphilosophie. Vorlesungen und Vorträge 1929-1941. In *Nachgelassene Manuskripte und Texte*. Hg. von Klaus Christian Köhnke, John Michael Krois und Oswald Schwemmer, Bd. 5: Hamburg (ECN 5).

Möckel, Christian. 2005. *Das Urphänomen des Lebens*. Ernst Cassirers Lebensbegriff. (Cassirer-Forschungen Bd. 12) Hamburg.

Natorp, Paul. 1912. *Allgemeine Psychologie nach kritischer Methode*. Erstes Buch: Objekt und Methode der Psychologie. Tübingen.

Windelband, Wilhelm. 1919. *Präludien. Reden und Aufsätze zur Philosophie und ihrer Geschichte*. 6. Aufl., Erster Band, Tübingen.

Chapter 16
Appearance or Existence of the Entity Realism 'Sense' or Mind

A. Yazdani

Abstract Two limiting boundaries exist in the domain of physics: the prime task of fundamental physics is to understand the objects, laws, or whatever is the basis of all observed phenomena. The other is the existence of a rhythm and pattern between the phenomena of nature (not apparent to eye, but to the conscious mind of analyzer). Also the development of scientific techniques has increased our ability of detailed observation and so the complexity of formalism. Consequently the concept of basic conditions and the character of physical laws as the fundamental principles of physics should be defined. It should be thought that, how someone can find a formalism to understand the phenomena or/and conduct an experiment. What do the concepts mean? What are the phenomena and under what conditions they exist or could be considered? What are the correspondence principles of mind and phenomena as the phenomena is constructed of two components existence and appearance, where only one of them can be measured? Since the cause of connection or/and interaction of both the observer and phenomena is space, its nature and related functions (equations) and formalisms should be considered. However, suppose we were able to find a theory which explains all observed (or existing) phenomena but, what are the basic processes through which, one can relate a behavior and character to the essential concept of a phenomena? There should be a conscious mind to be able to speculate the relation between basic conditions, physical parameters and even physical principles.

Keywords Invariant phenomena • Symmetry space

A. Yazdani (✉)
Tarbiat Modares University, P.O.Box:14115-175, Tehran, Iran
e-mail: yazdania@modares.ac.ir

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology, and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_16,

16.1 Introduction

The progress of modern science concentrated its penetrations into the wide range of mysterious research in this century. Where from one point of view the physical phenomena should be defined and on the other hand the observer as cognitive architecture should be considered. Even though challenge is posed to connectionism and cognitive architecture (Fodor and Mc Laughlin 1990), the response to the challenge is still an open question where (i) the systematicity is a type of response to challenge of 'how connectionism could be explained the systematicity' (McLaughlin 1993a) and (ii) cognitive psychology and computerized brain-imaging technology has also led to a revolution in techniques for measuring cognitive process in human (Nature 1994). How ever since both, the phenomena and cognition of the observer are a part of the universe (nature) as one system, the correspondence principle is considered, based on the symmetry which could be an essential condition of systematicity. Based on the symmetrical universe, the exchange interaction of the mind and the phenomena is the main point of consideration (Fig. 16.1). Principles of symmetry played little explicit role in the theoretical physics. The everyday definition of symmetry is;

I. Having an exact correspondence in size or shape between opposite sides of a structure, or
II. The regularity between parts of exchange where the both sides look exactly the same.

The generalized crystallography closely related to the theory of symmetry applied with the success concepts of packing, order and structural hierarch to discuss the complex natural structures (Weyl 1952). The ever growing attention is attributed to a special type of the hierarchical ordering on the fractal structures and on the

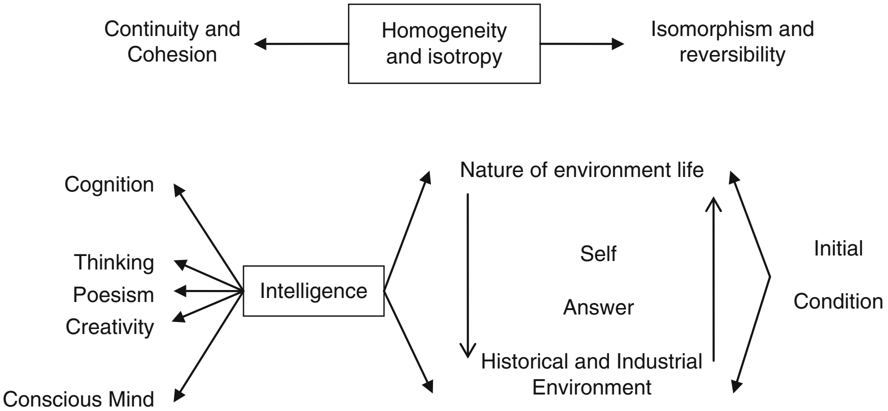


Fig. 16.1 The effect of initial condition "nature and society environment" on the reflectance of intelligence process.

other hand, the existence of some systematic relations among cognitive capacities (Mc Laughlin 1993b). In addition, the topological and symmetrical properties of the space and time, gauge invariance, act in physical space for conservation law of the surrounding world (Hill 1957).

Thus the symmetry, systematicity, regularity and the ordering parameters seem to be absolutely fascinating to human mind, as a conscious mind should be. This should be defined as the correspondence principle of structure of geometric of conscious mind and space. In this case anyone have a feeling or imagination about the physical laws of nature which is very close to the feeling of corresponding symmetry of objects and mind, namely the symmetry of the laws.

16.2 Concepts and Principles

The existence of the entity realism is a beauty of the truth which the natural phenomena could be sensed by someone. The importance of work style which is due to the genius as well as 'beauty of truth' is a much studied phenomena since the revolution of the twin theory of relativity and quantum mechanics. Conceptual truth, evolutions and regularities embodied in the laws of nature, would be able to make sense of physical events if the conscious mind of people could clear out the physical phenomena from the following points of view;

i. Initial condition of observer and strength of the conscious mind to formulize the experimental measurements,
ii. The initial conditions of the phenomena and more important the fundamental principles of nature, namely conservation laws, on which the phenomena can be considered as well as boundary condition, in the physical space (Figs. 16.1 and 16.2)

The ability to repeat experiments at different places and times is based on the invariance of the law of nature 'space-time' transformation. Therefore the fundamental physics is to understand;

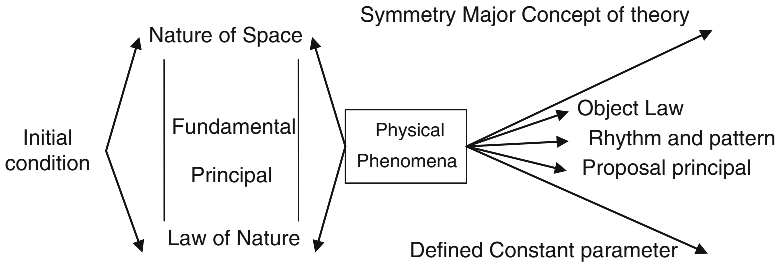


Fig. 16.2 Exchange of the space and the law of the nature on the objective and subjective

1. The objects laws which are the basis of all observed phenomena consider the ‘causes’ which depends on the perception of intelligent mind (thinking) (Fig. 16.2) which should be due to the symmetry of mind which is necessary for geometrical organized thought.
2. A rhythm and a pattern between the phenomena of nature which is not apparent to eye, but to the conscious mind of analyzer ‘systematically’. It should be persisted that, the principle of physics is due to the fact which says, various objects seeks to achieve order and harmony in their behavior, which is a characteristic of synchronization. This seems to be a manifestation of the nature tendency to self organization existing in nature. There are the laws of nature which summaries the regularities that are independent of the initial conditions. This is the entity realism of system which follows the ordering parameter.
3. And finally developments of scientific techniques or formalisms based on proposed principles (Constant principles) to discover and prove some sensible phenomena which are not observable. And it has increased our ability of detailed observation due to the strength of conscious mind.

However from each point of view the fundamental principles of physical laws are invariant to natural phenomena which are the initial principals for regularity and ability to repeat the observed measurements in different times and in different points of space, namely;

$$y(x, t_1) = y(x,t_2) = \ldots = y(x, t_n)$$

$$y(x_1, t) = y(x_2, t) = \ldots = y(x_n, t)$$

$$y_{o1}(x, t) = y_{o2}(x_2, t) = \ldots = y_{on}(x, t)$$

This means that the phenomenon is independent of the place (coordinate), time, and observer which are the best character of universality of physical laws. But how can we define the invariant principles? Winger argued that invariant principles provide a structure and coherency to the laws of nature just as the law of nature provides a structure and coherency to a set of events.

On the other hand how the connectionists can explain the systematicity? How they define the initial parameters and internal coherency corresponds to compositional structure if they avail themselves as distributed mental representations (Hargettai 1986/1989) (Fig. 16.2)? Anyway the existence of systematic relations between cognitive capacities and phenomena without assuming that mental processes are casually sensitive to constituent structure of mental representation is impossible. Thus ordering system, systematicity and symmetry for periodicity of patterns of space are inevitable. Indeed it is hard to imagine that any progress could have been made in deducing the laws of nature without existence of certain symmetry, which is the principle of equilibrium in universe, in presence of binary relations and geometrical interpretations in correspondence principles of mind and phenomena.

However the fundamental principle which governs the conservation of energy should be symmetric of space because of invariance of the phenomena. The primary concepts and initial principles on which the symmetry of space are defined are; continuity, homogeneity and isotropy of space. Related to the isotropy of space in the law of conservation of angular momentum, which is defined to be unchangeable to physical phenomena under rotation, with homogeneity is the conservation of linear momentum and impulse law which is defined the unchangeable phenomena under translation of the system. The translation (T) and rotation (R) (as symmetry group) are two important operators to understand the following concepts;

(a) The symmetrical objects on which the reflection and glide reflection should be added.
(b) The correspondence principle in topology.
(c) And finally the definition of motion which is space-time configuration.

The continuous, homogeny, and isotropic behaviors as well as linearity are the metric properties of space-time, which are the basic conditions needed for reversibility of the system according to the fundamental principles of physical laws. In this case the homogeneity of time which is related to the conservation of energy is another independent ordering parameter for symmetry of space.

Above all the defined factors of symmetry in space, the human brain (conscious mind)’s poesism is a recognizer of pattern and order, which possess a formalized beauty. Proposed principles, based on the corresponding life and consciousness, have emerged together in the universe with pattern and order. Thus the correspondence of appearance and existence of the phenomena as well as the conscious mind of observer in one system is suggested to make a definition for symmetry. Even though the geometrical symmetries are rather fundamental and any derivations must be somewhat speculative, appearing near the top of the quantum stair case, rotational symmetry and space-time translation symmetry are just other parts of the defeomorphism symmetry of generated relativity, which suggests their origin to be in gravitation theory.

However in spite of all definitions of symmetry on both points of view (i) the entropic principle implies that some symmetry has to be present because of environment and observers where the defined pattern means any regularity that can be recognized by the brain and (ii) the restrictions of mind to explain the symmetry as a consequence of the feature of the physical law, as much as possible in some restricted regime of physical phenomena. Let us take the beauty of fundamental geometrical strategies and structures behind the strength of mind of Newton - in his gravity law-, Coulomb – in the form of relation between electrical field and particle- and Biot and Savart – in magnetic field formula –in their formulas which all obey the inverse square law of action distance.

How the students explain the gravitation law? Why they immediately say that the force *between* objects? And varies inversely as the square of distance between them?

$$\vec{F} = \frac{GMm}{r^2}\widehat{r} \; or \; \vec{F} = \frac{KQq}{r^2}\widehat{r}$$

Why inverse of square distance? And what are the primary concepts of force between objects?

A specific form can gain a deeper insight into this modified formalism the Newton's third law and symmetry of space-time as a principle are involved. Consequently to reach the answer the essential condition of the inverse square law should be clear up. Since the defined physical space is the major and the least understood problem of its topological properties which can have; (1) the homogeneity and isotropy, which are the basis of symmetry, (2) continuity and cohesion, which are the causes of existing of the object defined by a slow, continuous and quasi static motion curve, can exist if and only if the function is irrotational as single value or monogenic.

Even though the Poincare invariance can be considered to be a consequence of general coordinate invariance in the absence of 'prior geometry' in gravitation theory, but the it is an essential condition for the geometry of space-time that is fixed immutably, and can be changed by changing the distribution of gravitation source (Misner et al. 1973). Therefore based on the above suggestion the universal law of inverse square radius (Williams et al. 1971) should be considered as follow;

(a) The central source of force, which is the 'cause', should effect but should not be affected from its environment. Thus it should be a point particle (charge or mass point) with mass of infinity ($M \to \infty$), named interval mass.

(b) The infinite mass of point particle should be located at central force around which all of the effected point objects have the unity mass or charge, and should not affect the central force ($m \to 0$). It should situate on the equipotential surface as homogenous space (defined as coherent object point), named spherical space. This is the character and the essential condition of stable equilibrium of space and is named the ground s-state in the atomic scale. The location of central force is where by increasing the effective distance (radius of sphere) the homogeneous field decreases proportional to the surface of sphere. And the effected object (m or $q \to 0$) is located on the surface of sphere of $4\pi r^2$. Therefore the following equations can be written;

$$\underset{M\to m}{\vec{F}} = \lim_{\substack{M\to\infty \\ m\to 0}} \left[\frac{M}{4\pi r^2}\right] m\hat{r}$$

$$\underset{Q\to q}{\vec{F}} = \lim_{\substack{M_Q\to\infty \\ q\to 0}} \left[\frac{Q}{4\pi r^2}\right] q\hat{r}$$

Where, M_Q is the mass of central charge and $\hat{r}$ is the unit vector in the direction of $\vec{r}$. There should be an important effective parameter which could be the character of both the filled material in the space of the central force and the effected object. In the electromagnetic case the character parameters are ε, the geoelectric permittivity, and μ, the geomagnetic permittivity, where $k = \frac{1}{4\pi\varepsilon}$ we have;

$$\underset{Q\to q}{\vec{F}} = \lim_{M_Q\to\infty}\left[\frac{Q}{(4\pi r^2)\,\varepsilon}\right]q\widehat{r} \to \lim\left[\frac{kQ}{r^2}\right]q\widehat{r}$$

Therefore in the gravitational force the equivalent parameter should be cleared up. By replacing G by the correspondence of the $K = 4\Pi\varepsilon$ where the ε should be property of the filled up the matter between the transfer the effect to the cause by which it should be clear in $G = \frac{1}{4\pi\varepsilon_g}$ (Aharoni et al. 1972) on which the minus sign insures that the gravity is attractive we would have;

$$\Rightarrow \underset{M\to m}{\vec{F}} = \lim_{\substack{M\to\infty \\ m\to 0}}\left[\frac{GM}{r^2}\right]m\widehat{r}$$

In this case if the mass of object, 'm' is equal to unit the resulting field of it is can be defined by:

$$\vec{E} = \lim_{m\to 0}\frac{\vec{F}_{M\to m}}{m} \to \vec{E} = \frac{GM}{r^2}\widehat{r}$$

Thus the interpretation of this formalism can be defined only in terms of central 'cause' in the uniformly spherical coordinate system on which the s-state can be compared with the Aristotle's 'God', as it could effect without being affected, even without any knowledge of its effects. Consequently if the inverse square of effective distance is accepted, the curvature of space which is due to the central force of s-state should be acceptable too (Williams et al. 1971).

On the other hand if the central force is located on the point particle 'm' while we suppose the distance of two objects is large enough, (compare to the radius of objects) we would have;

$$\underset{m\to M}{\vec{F}} = \lim_{\substack{m\to\infty \\ M\to 0}}\left[\frac{GM}{r^2}\right]M\widehat{r}$$

Therefore if the two forces are equal, since local inhomogenities have no observable effects on the inertial mass then it is well be acceptable that the global structure of the universality of $\vec{F} = \frac{GMm}{r^2}\widehat{r}$ exists between the objects if (a) the system of two objects is in equilibrium and (b) the second law of Newton is governed.

However the suggested conditions, on which the central 'cause' is completely at rest (named inertial mass M $\to \infty$), are proper if and only if the central inertial frame of reference is located on the central cause. Two points of view can be suggested here; (1) if the third law of Newton is governing in the whole system of cause and effect which is in stable state, the curvature of space cannot be considered where both the reference frames, observer and the central cause, are located at the central force but (2) if inertial frame of reference is not located on the central cause, and thus the effect of curvature can be considered regarding to the motion of particle point in the space-time of Minkofski (R4). It should be mentioned that in both cases a coincidence of two symmetries of space, natural phenomena and the conscious mind of observer.

References

Aharoni, A. 1972. *Lecture on mechanics*. Oxford, London, 310.

Fodor, Jerry, and Brain P. Mc Laughlin. 1990. *Cognition* 35: 183.

Hargettai, J. 1986. *Symmetry 1: Unifying human understanding*. Pergamon: Oxford and *Symmetry 2*, 1989.

Hill, E.L. 1957. *Review of Modern Physics* 23: 253.

McLaughlin, B.P. 1993a. Systematicity conceptual truth, and evolution. *Royal Institute of Philosophy Supplement* 34: 217–234.

Mc Laughlin, P.B. 1993b. *Philosophical Studies* 70: 45.

Parricia S. Goldman-Rakic 1994. The measure of the mind. *Nature* 369 (30 June): 717

Misner, C.W., K.S. Thorne, and J.A. Wheeler. 1973. *Gravitation*. San Francisco: W. H. Freeman.

Weyl, H. 1952. *Symmetry*. Princeton: Princeton University press.

Williams, E.R., J.E. Faller, and H.A. Hill. 1971. *Physical Review Letters* 26: 721.

Chapter 17
Fiction, Counterfactuals: The Challenge for Logic

Brian Hill

Abstract Fiction poses a set of problems relating to the notion of fictional truth and the reading of a fictional text, which link in with questions regarding counterfactuals and belief change. They are thus problems which modern logic, in this instance nonmonotonic logic, would like, or should like, to deal with. However, logics have, so far, not had much to offer. This paper argues that this is because logicians are not treating the correct problem. They assume in their logics an exogenous factor (similarity between worlds, epistemic entrenchment, selection function, to take but a few examples), whereas what is required is an understanding of this factor and the value it takes. This question essentially involves the dynamics: understanding this factor amounts to giving an account of how it comes to have a particular value in a particular context or at a particular moment. The paper contains a consideration of what should count as the essential tasks of such a theory of dynamics, and a tentative suggestion as to one direction for developing such a theory.

Keywords Belief revision • Counterfactuals • Dynamics • Reading of fiction • Similarity between possible worlds • Truth in fiction

There is a thorny and delicious problem, concerning the analysis of what it is for a sentence to be "true" according to a fictional text, which has enthused philosophers and logicians alike. Literary theorists are generally unimpressed by this problem, relying on notions such as "encyclopaedia" or "system of thought" to determine which sentences are true according to a fictional text.

There is a thorny and delicious problem, regarding the analysis of counterfactual conditionals (such as "if this match were struck, it would light"), which has enthused philosophers and logicians alike. Linguists, along with many language

---

B. Hill (✉)
HEC Paris/IHPST, Paris, France
e-mail: brian@brian-hill.org

O. Pombo et al. (eds.), *Special Sciences and the Unity of Science*, Logic, Epistemology, and the Unity of Science 24, DOI 10.1007/978-94-007-2030-5_17,

users, are generally unimpressed by this sort of problem, relying on notions such as "background assumptions", "conversational context" or "pragmatic factors" to deal with counterfactual conditionals and the role they play in everyday and scientific reasoning.

Logicians and logically-oriented philosophers will, in each of the above cases, balk at those who fail to be impressed, pointing at the lack of precision and clarity in the notions they are using in their approach to these phenomena (system of thought, conversational context). This despite the fact that these notions are the terms of art in the respective domains. Evidently, logically-oriented philosophers and logicians say, they can do better. Have they?

They have not; this paper shall be concerned with understanding why not. The explanation does not imply that they cannot, but only that they cannot if they continue doing what they are doing. Their failure, it shall be argued, lies in the fact that they have not quite grasped the problem they should be trying to solve.

The re-definition of the goal of such logical investigations will have consequences for what logic should be, at least if it wants to be of use to those involved in understanding and using the likes of fiction and counterfactual conditionals (henceforth called just counterfactuals). Several new demands on logic shall be formulated, and some brief suggestions for a framework which satisfies these demands shall be made.

17.1 Introduction: Bringing Logic to Fiction

Fiction, counterfactuals: two subjects which at first sight have little to do with each other. However, as is evident from the work of the last 30 years, the relationship between them is deep and important. A first link is most famously established by David Lewis (1983). He proposed an analysis of "fictional truth" – the sort of "truth" which is possessed by some sentences which are strictly false (or without truth value), on the basis of their relationship to certain works of fiction. "Sherlock Holmes is a detective", although not strictly true (in so far as Holmes does not exist), is *true according to the Holmes novels*, in a way that "Sherlock Holmes is a baker" is not. Lewis proposed that fictional truth was to be analysed using counterfactuals. In fact, he proposed an analysis which goes more or less as follows:

> A sentence of the form "in the fiction f, ϕ" is true if and only if: *if* f were told as known fact, *then* ϕ would be true.

The analysans (the part after the colon) is a counterfactual conditional: that is, roughly speaking, a conditional where the antecedent is (assumed to be) false. Such conditionals play important roles in many aspects of everyday life, science, and philosophy, not least the philosophy of science. For example, they are heavily involved in debates on causality and induction.[1] As one might expect given this

---

[1] Two classic references are, respectively, Lewis (1973b) and Goodman (1954).

importance, they pose a deep, difficult and interesting challenge for philosophers. However, there is relatively little indication that Lewis intended his analysis as an affirmation of a connection between the *philosophical* properties of counterfactuals and those of fiction (for example, there is no suggestion that they play similar roles in considerations of phenomena such as causality or induction). Rather, the analysis highlights the *logical* affinity between fiction and counterfactuals: truth in fiction and the truth of counterfactuals are related and obey similar laws. However, it is well-known that counterfactual conditionals, whatever roles they may play and problems they may pose in philosophy or linguistics, are appropriate objects for logical analysis: indeed, several philosophers and logicians, including Lewis himself, have developed "logics of counterfactuals" detailing the rules pertaining to the truth (or acceptability[2]) of counterfactuals and the consequences which can be drawn involving them. Lewis's theory of fiction in terms of counterfactuals implies that such rules apply, in an appropriately translated form, to fictional truth; to this extent, his analysis brings logic to fiction and indeed, is one of the first and most important theories to do so.[3]

This paper will be mainly concerned with evaluating the logical theories or systems in so far as they can be applied to fiction and counterfactuals. Before embarking on the discussion, it is worth pointing out the extent of the connection between fiction and the work of (some) logicians. Indeed, although Lewis's analysis in terms of counterfactuals has played a major role in bringing logic to fiction, the relationship between the two is perhaps deeper than the analysis might suggest. On the one hand, it can be established in other ways, without involving counterfactuals; on the other hand, it goes beyond the question of fictional truth.

The "branch" of logic pertinent for fiction is *nonmonotonic logic*. A nonmonotonic logic is a logic in which consequence is not necessarily preserved under strengthening of the antecedent: in such a logic, if 'C' is a consequence of 'A', it does not follow that 'C' is a consequence of 'A and B'. It has long been known that the logic of counterfactuals is a nonmonotonic logic, in the sense that, by adding a conjunct to the antecedent of true counterfactual, one may obtain a counterfactual which is false: "if this match were struck, it would light" is true, whereas "if this match were stuck and there were no oxygen in the atmosphere, it would light" is not. Lewis's analysis of fiction in terms of counterfactuals, in so far as it connects the logic of counterfactuals to the logic of fictional truth, thus establishes a relationship between fiction and nonmonotonic logic. However the detour via counterfactuals is strictly speaking unnecessary: this relationship can be defended on independent grounds. It suffices to note that fictional truth displays nonmonotonic behaviour with respect to parts of a work, or sequels of works: "in the first half of the detective story,

---

[2] Some, such as Adams (1975), have suggested that one cannot talk of truth or falsity of conditionals in general, and counterfactuals in particular, but only of acceptability; the logic of counterfactuals would thus aim to capture the rules relating acceptable counterfactuals. In this paper, these debates shall be ignored, and the discussion, where relevant, will be formulated in terms of truth.

[3] Another notable attempt is that of Woods (1974), although he does not propose such a complete analysis of fictional truth.

the butler was the most likely culprit, in the whole work, it was the gardener"; "in the original Star Wars film, Luke Skywalker and Princess Lea were just friends, in the whole trilogy, they were brother and sister". Once this is noticed, the application of nonmonotonic logic to fictional truth comes to seem natural.

But the points of contact between fiction and (nonmonotonic) logic are by no means limited to the notion of fictional truth: notably, the important question of the beliefs the reader comes to have as regards the fictional text may also lend itself to a treatment with the tools of logic. For some philosophers of fiction, the question of the reader's beliefs as regards the text and the question of fictional truth are very closely related: they reject the role of the narrator in Lewis's analysis ("... if f were *told* as known fact ..."), formulating analyses that rely on the reader rather than the narrator or the author.[4] The general form of such analyses is as follows:

> A sentence of the form "in the fiction f, ϕ" is true if and only if: if the reader were to come across f in circumstances which led him to believe it to be known fact, then it would be reasonable for him to believe ϕ.

In other words, fictional truth is related to the reader's (real) beliefs in the (counterfactual) situation where he believes the text to be true; in particular, it is related to the *new beliefs* which the reader *acquires* on reading the text (in this counterfactual situation). It is true in the Holmes stories that Holmes is a detective if, had the reader believed the stories to be known fact, he would, on reading them, have acquired the belief that Holmes is a detective. Whether or not one accepts this analysis of fictional truth, there is undeniably a close relationship, and indeed one which is important for the study of fiction and of the interaction between human beings and works of fiction, between the beliefs a broadly competent reader comes to have *about* the fictional text (in the actual world, where he knows it to be false), and the fictional truths of the text. A competent reader of Conan-Doyle comes to believe that, according to the Holmes novels, Holmes is a detective, and indeed, this is a fictional truth of the Holmes novels. A central question for theorists of fiction (be they philosophers or literary theorists) is thus that of the changes in the reader's beliefs (and specifically, his beliefs regarding the text) during the reading process. One important account of belief change in the face of new information (for example, in the process of reading) is given by theories of belief revision proposed by philosophers, logicians and artificial intelligence scientists. As has been established by the work of Gardenförs, Makinson and others,[5] the logic of belief revision is nonmonotonic. There thus appears to be a deep link between several of the questions regarding fiction – pertaining not only to fictional truth, but also to belief about fiction – and nonmonotonic logic.

It is no surprise that counterfactual conditionals lend themselves to logical analysis. It should now be equally clear that fiction, and more particularly questions such as those regarding what is true according to a fictional text and what the

---

[4] See Matravers (1995) or Krasner (2002).

[5] See for example Gardenförs (1988), §4.5 and Ch. 7, and Makinson and Gardenförs (1991), for example.

reader learns from a fictional text, is apt for the application of logical methods. It seems no exaggeration to think of fiction and counterfactuals as examples of the potential pertinence of logical methods to questions of particular disciplines, such as psychology (of reading), literary theory (fictional truth), linguistics (of counterfactuals) and philosophy (of fiction, of counterfactuals and of related questions).

17.2 Problem: Confronting Logic with Fiction

However, the contribution made by logic to the questions raised above is at best limited, at worst warped. A brief examination of Lewis's full analysis of fictional truth will make clear why this is the case.

Lewis's interest in truth and fiction, and in the analysis in terms of counterfactuals, is partially related to the possibility of applying his own, famous, theory of counterfactuals to fiction. Roughly, for Lewis, a counterfactual "if ψ, then ϕ" is true if and only if, for any possible world where ψ is true and ϕ is false, there is a possible world which is *more similar* to the actual world were both ψ and ϕ are true (Lewis 1973a). Applying this theory of counterfactuals to his analysis of fictional truth, he obtains:

> A sentence of the form "in the fiction *f*, ϕ" is non-vacuously true if and only if some world where *f* is told as known fact and ϕ is true differs less from our actual world, on balance, than does any world where *f* is told as known fact and ϕ is not true. It is vacuously true if and only if there are no possible worlds where *f* is true as known fact. (Lewis 1983, p. 270)

If this analysis has had any influence beyond the restricted circle of philosophers concerned with proposing analyses of the notion of fictional truth, it is fair to say it has only been in the form of "secondary consequences". Lewis's article figures alongside other classic works of analytic philosophers and logicians (such as Kripke's original articles on modal logic) as one of the sources of the notion of "possible world", which many literary theorists have taken to heart.[6] Literary theorists have adopted notions such as this one, which are at least partially logical, perhaps without remaining entirely faithful to the logical source: thus Eco (1985, p. 158) opts for the notion of "furnished worlds" ("not abstract types of possible worlds [. . .] but 'pregnant' worlds of which one must know all the acting individuals and their properties") in his analysis of the reader's appreciation of the text; Pavel (1981, p. 174) develops the notion of "cosmos" ("an ordered triple containing a nonnull set C of possible worlds, a world W belonging to C and given as the actual world and a binary relation A on C, the relation of accessibility"), using the comparison of different "cosmoses" to analyse the "multi-level ontologies"

---

[6] "Literary theorists [. . .] have developed a textual semantics based on the idea that the semantic domain projected by the literary text is a non-actual possible or an alternative possible world", Ryan (1991), p. 553.

involved in fiction; and Ryan (1991) employs different “types of accessibility relations” (which closely resemble Lewis’s notion of similarity between worlds) to propose a characterisation of literary genres. If logic was to have something to offer to other domains of investigation, it was rigour and power, probably not this sort of “inspiration”.

As well as the notion of possible world, Lewis’s analysis involves another notion: that of similarity between worlds. This notion plays the central role in his analysis of counterfactuals, even though, as many have pointed out, and as he himself was quick to emphasise, no deep understanding has been proposed. Doubtless we seem to have an intuitive sense of similarity, and perhaps of similarity between worlds, and certainly this adds to the appeal of any analysis which employs the notion. Nevertheless, the notion is vague, variable, and the chances of a rigorous analysis are slim. As Lewis puts it, quoting Goodman, similarity “is a highly volatile matter, varying with every shift of context and interest”.[7]

The reliance on such an elusive, if intuitive, “factor”, which must be assumed as primitive, is by no means a particularity of Lewis’s theory. On the contrary, any theory of counterfactuals, and indeed any nonmontonic logic (thus logics of belief revision) involves such a factor, although the names and forms given to it may vary. Theories of counterfactuals which do not use a relation of similarity between worlds have to suppose some “relevant conditions” and “connecting principles” (Goodman 1954), a variably interpreted modal operator (Lowe 1995), or an appropriate moment in the past where the antecedent of the counterfactual is still realisable (Adams 1975; Edgington 1995; Placek and Müller 2007). More generally, nonmonotonic logics require an extra structure, whether this be an ordering on states (comparable to the similarity relation on worlds), a selection function yielding maximal sets of formula containing the premise or antecedent, an order of “epistemic entrenchment” on propositions, or a set of default propositions, to take but a few examples (Makinson 2005; Gardenförs 1988). In many of these cases, logicians and philosophers admit that, in real applications of the logical machinery, the relevant factor is liable to fluctuate between different moments or contexts.[8]

In some cases it may well be that the factor employed in the theory is at least more comprehensible than the counterfactual, non-monotonic consequence or change of belief to be understood, in the sense that it can be fixed independently of the decision as to the truth of the counterfactual, the validity of the consequence or the adequacy of the change. The application of default logic to cases where the defaults are known is such an example. However, in a large number of cases considered by philosophers

---

[7] Lewis (1973a), p. 92. The original quote is from Goodman (1970).

[8] This is the case for Goodman (in so far as the connecting principles are laws and depend on an epistemic entrenchment ordering which may alter through time) and Lowe (p. 49) in the case of counterfactuals; in belief revision, theorists working on iterated revision face the question of changes in the factor (see for example Rott (1999, 2003) for a technical and philosophical discussion).

and logicians, such as the questions of counterfactuals and fiction at issue here, this is not the case. There is no independent way, bar simple intuition, of deciding which worlds are most similar to the actual world, or deciding which selection function or order of epistemic entrenchment to use in belief change (and in particular, in the changes of belief which occur in the reading of fictional texts). Indeed, according to Lewis, it is the other way round:

> The thing to do is not to start by deciding, once and for all, what we think about similarity of worlds, so that we can afterwards use these decisions to test Analysis 2 [his analysis of counterfactuals]. … Rather, we must use what we know about the truth and falsity of counterfactuals to see if we can find some sort of similarity relation – not necessarily the first one that springs to mind – that combines with Analysis 2 to yield the proper truth conditions. (Lewis 1979, p. 466–7)

This has two significant consequences for the analyses which these logics can offer of subjects such as counterfactuals, fictional truth and belief about fiction.

Firstly, it follows that such analyses will be of only limited help to those who *use* or are *interested in* the analysandum. If in doubt, such analyses of counterfactuals will not generally tell you which counterfactuals are true. If there is debate, such analyses will not be able to decide on the fictional truths of a work. If desired, such theories of belief revision will not shed much light on the processes involved in the reading of fictional texts, nor in the results of these processes. Their advantage, if any, is at a more general level: they will perhaps furnish general properties of counterfactuals, fictional truths, or the reading of fiction, properties which correspond to the validities of the logical systems. Indeed, as Lewis points out, this is basically the only sign of the correctness of the theories themselves: his theory is "not devoid of testable content – it settles some questions of logic" (Lewis 1979, p. 465).

However, it turns out that even this meagre set of logical properties cannot constitute a test of these sorts of theories. In the case of counterfactuals, different theories of counterfactuals, involving different properties of the similarity relation between worlds, and thus yielding a different set of logical validities, can all be defended by operating judicious choices of the similarity relation. In a detailed discussion of different theories of counterfactuals, imposing different conditions on the similarity relation between worlds and thus yielding different sets of logical validities, Lewis comes to the conclusion that the essential difference between them is whether they see the linguistic context as furnishing a determinate, single, similarity relation (with weak properties) or a class of similarity relations (each having strong properties).[9] Another illustration of the dependence on the similarity relation in the resolution of logical questions concerns the apparent counterexamples to logical rules which are validated by *all* theories of counterfactuals.The theorists'

---

[9] Lewis (1981). Stalnaker (1980) contains a defence of an analysis of counterfactuals which involves a similar dependence on the way the analysis is applied (ie. the way the similarity relations, or sets of similarity relations, are chosen).

reply typically consists in dissolving the counterexample by claiming that there has been a shift of context, with an accompanying, opportune, alteration in the similarity relation. This defence shall be discussed in more detail below.

These are two examples where the success or plausibility of the theories are strongly dependant on the intricacies of their application. They highlight the second point to be made about the theories proposed by these philosophers and nonmonotonic logicians: namely, the results of the application of the theories are highly sensitive to the choice of factor (in the example case, the choice of similarity relation between possible worlds). Accordingly, the efficacy of the theories suffer: if the application of a theory will be always controversial, especially when it comes to properties implied by the theory itself, then the theory becomes a lot less useful. Such a theory is more often than not to be avoided; it results in confusion rather than clarification.

The first point indicates that the nonmonotonic logics and the philosophical analyses which are on a par with them are generally unhelpful when it comes to understanding real phenomena, such as the phenomena of fiction (fictional truth, the reading of fiction) and indeed the everyday and scientific use of counterfactuals. The second point is more drastic: it suggests that the analyses are not only unhelpful, but practicable unusable.

17.3 Diagnosis: Crossing Disciplines

Before concluding that logic has very little to offer such investigations, before counting a point against the unity of science, understood as productive cross-fertilisation by the sharing of techniques – in this particular case, the techniques of logic – it is perhaps worth inquiring into the reasons behind the apparent unsuitability of the logicians' efforts to more concrete questions. If the hope is that logic can supply useful tools for application beyond the confines of logic itself, it is perhaps worth reflecting on the conditions under which cross-disciplinary applications are fruitful. The reflection will be brief; its goal is to draw attention to what logic *should do*, if it is to make a useful contribution to the sort of questions posed above.

Fruitful cross-disciplinarity does not require that the two investigations (or disciplines) involve the same sorts of objects. Applications of techniques developed in or for physics, say dynamical systems theory, to questions of biology attest to this, rather evident fact, in the sense that on the one hand one is considering systems with physical objects (for example, fluid flows), and on the other hand, there are biological objects involved (for example, populations). Furthermore, this example illustrates a type of cross-disciplinary exchange, which has two pertinent aspects. Firstly, although the ontologies of the two disciplines are different, there is a structural similarity between their objects of study (the systems). It is difficult, if not impossible, to specify in a non-circular manner what this means, but an intuitive understanding shall be sufficient for current purposes: they both are treating systems which, roughly speaking, have several general properties in common (for example, a "low level", with many actors, a "higher level" with relatively few parameters of particular interest). Secondly, both disciplines want to know approximately the same sort of thing about their respective systems: how they change in time and the nature of the relation between the low and high level, for example. In this sense, they may be seen as having a common aim, or at least aims of a common sort (understand such systems), whilst, evidently, the more specific aims of the investigations (understand fluid flow, understand population dynamics) differ, by the fact that they involve different realisations of similar structures.

Obviously, this is not the only instance of fruitful cross-disciplinarity, nor the only type of cross-disciplinarity. For the purposes of contrast, note that there seem to be cases of interdisciplinary reference which rely on a (partial) sharing of ontologies, despite a difference of aims (no matter the degree of abstractness at which the aims are described). Consider for example the use of number theory in coding theory (and more specifically, cryptography). There is an overlap in the objects of these theories (both use and consider numbers). On the other hand, the aims of the theories are notably different: one wishes (generally speaking) to understand the structure of the natural numbers; the other wishes to construct transformations from natural numbers to themselves which are such the inverse transformation is easy to find given a certain amount of minimal information, and as difficult as possible to find without it (i.e. find a code that is as difficult as possible to break).[10] The fruitfulness of the cross-disciplinarity rests on the fact that certain *results* of one of the theories (viz. theorems of number theory) are useful to the *aims* of the other theory (viz. constructing coding algorithms). The RSA coding algorithm, for example, rests on number theoretic results regarding the factorisation of products of prime numbers. Note that the sort of structural similarity involved in the previous case is less important here: it is rather the fact that the same sorts of objects are involved, and therefore that the results proved in one domain still hold in the other, which is crucial.

These examples illustrate two "paradigmatic types" of cross-disciplinarity; cases of cross-disciplinarity will generally display a complex mix of the properties involved in the above examples. The distinction between these types is sufficient, however, to allow a preliminary glance at the nature of the cross-disciplinarity one might expect from logic, in its application to questions regarding fiction, the reading of fiction, and counterfactuals.

Consider the possibility of using the techniques of nonmonotonic logics in investigations regarding fiction and counterfactuals. As noted above, the structures involved have a lot in common: fiction and counterfactuals display the sort of nonmonotonic behaviour which lends them to analysis with the techniques of nonmonotonic logic. One of the conditions for the former sort of cross-disciplinarity

---

[10] This is a very rough characterisation of both disciplines (both now involve structures other than the natural numbers, coding theory now takes more account of information theory, communication theory etc.), but it is sufficient for the point made here.

is satisfied. The other condition is that nonmonotonic logicians and those interested in fiction or counterfactuals share the same aim. A major aim of literary theory is to understand how the text makes the reader arrive at his conclusions, and more generally how the fiction influences the opinions of its reader (and thus the society to which he belongs). A major aim of the psychology of fiction is to understand how the text is treated, the role of the reader's beliefs and dispositions in this treatment, and the effect of this process on his beliefs and dispositions. One of the most difficult and important questions regarding counterfactuals concerns the way in which they come to be accepted or true, the effect of the fact of being accepted, and how, at other moments or in other contexts, they may fail to be true. In each of these cases, the question is not so much that of knowing which *are* the fictional truths, given a similarity ordering on worlds, or what *are* the appropriate belief changes, given an epistemic entrenchment ordering, or which *are* the true counterfactuals, given a notion of similarity between worlds. It is more like: how does the similarity relation *change* for different texts and different parts of the text, how does the epistemic entrenchment ordering *evolve* during and after reading, how does the notion of similarity *shift* from one moment to another and one context to another? If the logical analysis of counterfactuals, fiction, or belief revision relies on a factor, the concrete investigations into these phenomena need to know something about the *dynamics* of this factor. Traditionally, however, logicians have been largely content with assuming the factor as given, and investigating the properties of the logics relying on that factor, or comparing the logics which rely on different sorts of factor.

It is as if the aims of logic are not similar enough to the concerns of the relevant domains to allow fruitful application. This would explain the lack of cross-disciplinarity of the first sort discussed above. But it is equally true that there is no cross-disciplinarity of the second sort. Accepting for the sake of argument that logic deals with the type of object involved in the questions of fiction, of attitudes toward fictional texts, of counterfactuals (whether they be propositions, sentences, statements or utterances), it nevertheless remains that logicians' results are of little interest outside their domain. Few outside of logic have had reason to appeal to representation theorems[11] showing that, for example, the logic of counterfactuals is precisely rendered by the model in terms of similarity of possible worlds. Indeed, as noted above, the main use of possible worlds in the theory of fiction owes more to the ideas evoked by the notion of possible world that any logician's result.

The situation described in the opening paragraphs of this paper can now be better understood. The "encyclopaedia" and "background assumptions" of theorists of fiction, of mental attitudes to fictional texts, and of counterfactuals are the equivalents of the logicians' and philosophers' similarity between possible worlds and epistemic entrenchment. When the latter are taken as given, results about the ensuing logical structure can be proved; employing the former, on the other hand, it is hoped that a deeper understating of the concrete phenomena, including notably

---

[11] As Makinson (2005), pp. 28–29 points out, representation theorems are the nonmonotonic equivalent of completeness theorems.

the dynamic phenomena, involving fiction and counterfactuals, can be attained. Logicians (and logically-minded philosophers) are unimpressed with theorists of fiction, of mental attitudes to fictional texts, and of counterfactuals, because these theorists are worried about elements that logicians, to date, only assume. Inversely, these theorists are not much concerned with the work of logicians because the latter have to suppose what the former would like to understand. *If* there is to be cross-disciplinary exchange, one side will have to give. *If* logic is to be applicable to these questions, it will have to be the logicians.

17.4 Constraints: Requirements on a Logic for Fiction

A logic which will be more useful for understanding fiction, the reading of fictional texts, and counterfactuals will thus have to be a "logic" which takes up the challenge of the dynamics of the logical framework. Such a logic will thus have to break with much past work, which assumes the appropriate logical framework as given. In this section, some of the consequences of this re-conception of the task of logic shall be considered, and some requirements on what such a logic would look like will be proposed; in the following section, a few suggestions as to possible directions for developing such a framework shall be proposed.

Given that, traditionally, logicians have not devoted much attention to the dynamics of their frameworks,[12] it does not seem unwise to take a glance at other domains where the question of dynamics has received a thorough treatment, such as physics. Consider the dynamics of a physical system (to fix ideas, consider a system of classical mechanics). At each moment, all of the bodies of the system have various values attached to them (for position, velocity and so on); from one moment to the next, these values change. This does not happen chaotically, but according to by now well-understood laws (the laws of physics). One important factor of these laws is that the change in many of the most important factors – position, velocity, for example – is continuous: that is, the value at one moment is close to the value at the next. Indeed, if it were not continuous, it would be practically impossible (at least for us, finite-minded humans) to understand the system. Suppose movement were not continuous, that particles jump about from one point in space to another from one moment to the next. Suppose furthermore that there were no regular pattern in this jumping about; one would say in this case that we cannot understand the

---

[12] By dynamics of the frameworks, we mean change in the machinery assumed for evaluating propositions; in the case discussed above, possible worlds and a similarity relation on them. Therefore, traditional forms of dynamic logic (van Benthem 1996) do not take account of the dynamics of their frameworks, in so far as they assume a set of states and relations on them without posing the question of possible changes in the set of states or the relation. Those who have begun to look at this question, at least to a certain degree, are rather theorists working in logics for knowledge change under communication (for example Baltag et al. 2004) and some who consider iterated belief revision (for example, Rott 2006).

system (and perhaps, even, there is no understanding to be had). Suppose on the other hand that there is a regular pattern: for example, that, for any particle that is at point p at one moment, it will be at one of the points in a set P at the next moment. Things would be a lot clearer (supposing that one's unique goal is to understand this movement), if one "re-arranged" the points so that the points in P are "beside" the point p.[13] This is apparently a first step towards a deeper comprehension of the system. Under such a re-arrangement – in such a (new) "space", on the same (old) points – the movement of the particles *is* continuous. Continuity is thus an attractive, and essential, property of system, in so far as the question of its dynamics is concerned. Furthermore, it is, at least partially, a property of the *way the system is represented*. In the re-arranged space, the movement is continuous; in the original space, the movement is not. That is to say: if each instantaneous state of the system is represented in the re-arranged space, then the dynamics of the system, understood as the sequence of instantaneous states, is continuous. Continuity is not just a property of the dynamics of the system, but equally a property of the framework used for representing the instantaneous states of the system.

This example illustrates two important aspects of the question posed by the dynamics. The first can be stated as a principle, the *continuity principle*: a theory of the dynamics (of some system) should represent the system as changing continuously. This is a methodological principle, advising how one could begin to tackle the question of the dynamics; it cannot be defended but by examples, such as the one above, of theories which have flourished once a representation allowing continuous dynamics was found.

The second moral of the example concerns a requirement which results from the continuity principle. There is a requirement on the representation of the *instantaneous state* of the system: it must be represented in such a way that a notion of *proximity* between states can be naturally defined, according to which the dynamics of the system is continuous.[14]

Any logic which purports to support or comprise a theory of the dynamics of the logical framework is subject to this extra constraint. In particular, it must permit (at least) a notion of proximity between frameworks relative to which the dynamics of the frameworks is continuous. That is: it must provide a representation of the instantaneous (logical) state such that the dynamics of these states is continuous.

To illustrate this requirement, consider once again the analysis of counterfactuals in terms of similarity of worlds. The standard analysis takes the set of possible worlds and the similarity relation on these worlds as given, and provides an evaluation of all the counterfactuals (in each world). However, as noted above,

---

[13] Put more technically, the suggestion is that a (new) topology be defined on the (old) set of points.

[14] The notion of proximity and continuity (as proximity of consecutive states) are left loosely specified here, although they correspond to basic notions from topology. In particular, although proximity can be thought of as a distance (ie. the space is a metric space), it need not be – containment or overlap is sufficient (ie. the space is a topological non metric space). See for example Sutherland (1975) for details.

the similarity relation is liable to change between contexts. Therefore, the logical structure proposed by the standard analysis – comprising the worlds, the relation and the evaluation of counterfactuals which rely on these – is really a representation of the *instantaneous state*: namely, the state of (logical) affairs in that context. According to the continuity principle, this representation should be judged on its ability to support a representation of the dynamics as continuous. It satisfies the requirement only if a relation of proximity between similarity relations (and sets of worlds) can be defined such that the changes of similarity relations between consecutive contexts are continuous with respect to this notion of proximity. Grossly put, it needs a way of deciding whether one similarity relation is *close* to another, such that, if similarity relation *a* is effective in one context, and similarity relation *b* is effective in the next context (the context at the next moment), then *a* is close to *b*.

To illustrate the force of this requirement, consider the following example.[15] Jones does not like getting wet, but had to go out. The following two counterfactuals must thus be true: "if it had rained (A), he would bring an umbrella (B)", and "if it had rained (A) and he had brought an umbrella (B), he would have been dry (C)". However, the following counterfactual does not seem to be true: "if it had rained (A), he would have been dry (C)". This contravenes the rule of cumulative transitivity, or cut, common to virtually all nonmonotonic logics (and all logics of counterfactuals), which allows the conclusion "if A, then C" from the premises "if A, then B" and "if A and B, then C".[16] To defend against this counterexample to their logical analyses, theorists of counterfactuals submit that there has been a change in the similarity relation between the moment or context where the premises are asserted (and are true) and that where the conclusion is asserted (and is false). When the premises are asserted, in the closest A-worlds (that is, worlds where A is true), B is true, and the closest A and B-worlds, C is true; so the closest A-worlds are all closest A and B-worlds, and so C is true in these. In this context or at this moment, the counterfactual "if it had rained (A), he would have been dry (C)" is true, and the rule of cut is valid. However, the defence continues, when one actually comes to evaluate the counterfactual "if it had rained, he would have been dry", one does so at a subsequent moment, or one moves into another context, with a *different* similarity relation between worlds. According to it, the closest A-worlds no longer all satisfy B, and indeed, no longer satisfy C, so that the first premise and the conclusion are no longer true. Such a shift of similarity relation does allow the analysis to survive the apparent counterexample; however, if the ambition is to account for the *dynamics* of the framework, it cannot come for free. It needs to be explained how the similarity relation in the second moment or context is "close" to the similarity relation in the original moment or context: indeed, close enough to be able to think of this change (and others like it) as continuous. The continuity principle requires one to tame the wild thrashing around of worlds involved in such discussions.

---

[15] The example is structurally similar to that considered by Stalnaker (1984), p. 130 *sq.*, as is the reply relying on context-shift.

[16] Lewis (1973a), p. 35; Makinson (2005), p. 5.

These considerations might give one reason to be sceptical regarding the traditional theories of counterfactuals, construed as representations of instantaneous logical states (at particular moments or in particular contexts). How can one understand the relation between the similarity relation used to evaluate counterfactuals at one moment and the similarity relation at the next moment if so many worlds which previously were similar to the actual one become distant, and vice versa? If such an understanding is not to be had, the representation proposed is inadequate: theorising about counterfactuals in terms of similarity relations on possible worlds does not satisfy the continuity principle, and is thus not appropriate for a full theory of counterfactuals, which seeks to understand, not only the "instantaneous" logical properties, but equally the changes from one moment or context to the next.

Whether or not one decides to reject the traditional theories on these grounds or attempt to salvage them, the example shows how the continuity principle counts as an extra constraint on a would-be theory of counterfactuals. As such, it should be particularly welcome given the aforementioned problems with the power and applicability of traditional theories. Recall that these theories are normally difficult to apply, not only because they employ a factor which is exogenous to the theory, and difficult to fix independently of it, but moreover because the results of the theory are very sensitive to the value of these factors. As the example above shows, even counterexamples to the most basic consequences of the theories (viz. logical properties of counterfactuals) can be evaded by a judicious choice of factor (similarity relation). The continuity principle places the onus on the theorist, asking him to justify (at least minimally) his choice of factor and the shifts in the factor. This means that some shifts will be permitted (according to the theory) and others will not: the wide range of similarity relations which could be effective in any given context or at any given moment will be thus restricted. Accordingly, this will mitigate the sensitivity of the results of the theory to such choices. Since it respects more constraints, the theory is stronger, and more "testable". The important point for current purposes is that it equally becomes more stable, more operable, and thus more interesting as a tool for understanding fiction and counterfactuals.

Furthermore, the dynamic constraints placed by the continuity principle may help with the first difficulty of traditional theories, namely the fixing of the relevant factors (similarity relation or epistemic entrenchment, for example). Consider once again the analogy with mechanics: nobody demands of a theory of mechanics that it determines the properties of the instantaneous state at a particular moment, solely on the basis of the state at that moment itself. There is no way of fixing the velocities or accelerations of the particles at t (if they were not given) on the basis solely of their positions at t. However, once the states at other moments are available, it becomes possible to fix the properties at each moment, by appealing to other moments, and by relying on the constraints imposed by the dynamics (of which the most basic is continuity). This often works in one of two ways: if one can suppose an original instantaneous state to be more or less specified, many properties of subsequent states can be more or less precisely determined. If, on the other hand, only certain information about states is available (for example, positions of particles at consecutive moments), then other information (velocity or force at different moments) can be more or less precisely determined. Such techniques may also be available to a logic which accounts for the dynamics of its factors. For example, if one can suppose that the similarity relation (to continue with the example) has a certain set of properties in one context, this may allow one to deduce that the similarity relations in subsequent contexts will have certain properties. Or, on the basis of varying opinions regarding certain counterfactuals in consecutive contexts, it may be possible to deduce the evaluations of other counterfactuals, using restrictions which the known counterfactuals, in tandem with the dynamic constraints, place on the similarity relations.

The case of counterfactuals, and in particular the analysis in terms of similarity of worlds, has been chosen because it is intuitive and well-known; the main points made above apply to the other questions discussed, and the other logical frameworks employed. In models of fictional truth relying on counterfactuals, say, one expects that a theory represents fictional truth in such a way that changes – on reading the sequel of the book, or following modifications in the opinions of the society where it is read, for example – can be accounted for. In models of reading, and logics of belief revision, the changes in belief and the learning of information should be represented in such a way that development in the mechanisms of belief change – that is, development of the dispositions to give up belief in X before giving up belief in Y – can be understood continuously.

Indeed, as for the case of similarity of worlds, it is possible to go through all the versions of nonmonotonic logic, and all their applications, asking to what extent they satisfy the requirements stated above. Such an evaluation shall not be undertaken here. Instead, the paper will close with some general, tentative remarks about properties of frameworks which could possibly help satisfy the continuity principle. A full defence of all the aspects would exceed the ambitions of this volume, and would require a more thorough discussion of existing theories.

17.5 Prospects: Towards a Logic for Fiction

Like many logical theories, theories of counterfactuals (and indeed, of belief revision) are often presented by first posing a question (what makes the counterfactual "if A, then B" true?), then stating an intuition (the fact that, in situations most similar to the actual one where A is true, B is also true), and finally going on to formalise it. If one is concerned with developing a framework which satisfies the continuity principles, perhaps a relevant question to ask is: what sorts of things *do* vary continuously from one moment, or context, to the next?

It certainly does not seem that the sentences or propositions accepted or evaluated as true in a context vary continuously: one can suddenly learn a sentence to be true or false. By contrast, it does appear that the collection of sentences or propositions *involved* at a given moment or in a given context varies continuously. Obviously, new issues may come into play between one context and the next; however, the latter context still retains a certain number of the propositions or sentences in play in the former. In this sense, one seems to be able to talk of continuity: between successive contexts, the sets of propositions or sentences *involved* or *in play* overlap.

The notion of a sentence or proposition being in play is certainly relevant to the questions regarding fiction and counterfactuals raised above. When considering fictional truths, there are certainly a certain number of questions which never (except perhaps in the odd philosophy paper) appear in discussion (for example, is it true in the Holmes stories that Gladstone was born in 1809?). Even more obviously, in the reading of fiction, at any moment there are issues which are not in play: indeed, many novels take advantage of the fact that propositions are not in play at certain moments of the reading (all the clues as to the perpetrator of the crime have been mentioned somewhere in the book, but rarely can the reader recall them all at one point and put them all together in the appropriate way). As noted above, for Lewis and many others, the similarity relation appearing in the analysis of counterfactuals depends on "context and practice"[17]; in any given context, however, there seem to be many sentences or propositions which are out of play, so that the notion of being out of play may prove equally relevant for theories of counterfactuals. When one considers the truth of "if this match had been struck, it would have lit", the outcome of the 2004 US presidential election is hardly involved.

However, it is generally recognised that the notion of sentences or propositions being in or out of play poses difficulties for most modern logical systems. The situation is most discussed, and most evident, in logics of knowledge and belief. Classical models (Hintikka 1962) suffer from problems of "logical omniscience" – they attribute to the agent knowledge of all logical truths and all logical consequences, knowledge which real agents seem not to possess. One of the important ways in which humans fail to actually attain logical omniscience is due to the phenomenon of awareness: they are only aware of a certain number of issues, or sentences, or propositions, and in particular they are not (always) aware of the logical truths or logical consequences. In other words, some sentences or propositions which are logical truths are *out of play* for such agents (at such moments). For example, the fact that $37{,}147 \times 5{,}643 = 209{,}620{,}521$ is probably out of play for most agents (at most moments). Being unaware of these sentences, they do not believe them to be true (or false, for that matter): thus the failure of logical omniscience. Notoriously, any system which assumes a fixed language equipped with a (classical) logical structure and models belief or knowledge in such a way that they are closed under logical consequence suffers from the problem of logical omniscience.[18] Examples of such systems not only include traditional modal logics for knowledge, which assume a fixed set of possible worlds interpreting the sentences of a given language and model knowledge by appropriate sets of worlds, but equally the major models for belief revision, which traditionally have worked with a fixed set of sentences closed under Boolean operations with a classical notion of logical consequence

---

[17] Lewis (1981), p. 87.

[18] See for example Modica and Rustichini (1994).

on them, modelling belief as sets of sentences closed under logical consequence (Gardenförs 1988).

Indeed, virtually all systems of nonmonotonic logic, in so far as they "piggyback" on classical logic (Makinson 2005) exhibit logical omniscience-like behaviour; in other words, they fail to capture the notion of being in play. This must be considered as a disadvantage of the models, because the importance of the notion of being in play extends beyond the questions of knowledge, belief and belief change. Consider conversational contexts for example: in a particular context, a certain sentence may not be in play (and thus the interlocutors cannot or do not assume it to be presupposed). However, if one models the context as a set of propositions closed under (classical) logical consequence – as a set of possible worlds, say[19] – the fact that the sentence is out of play is lost. For any sentence of the language, either it is true (respectively false) in all the possible worlds of the context, and so is presupposed to be true (respectively false) in the context, or it is true in some worlds and not in others, in which case it is "open question" in the context, suitable for debate and discussion: neither of these cases captures what it is for the sentence to be out of play. As discussed above, theories of counterfactuals typically rely on representations of contexts which are "extensions" of the notion described above, in so far as they consist not only of a set of worlds, but equally of a similarity relation on them. It follows that these theories cannot capture the notion of in play: all the sentences of the language are in play in every context. This fact has some peculiar consequences for their evaluations of counterfactuals. Consider a context in which the eventuality where a certain match is struck is being considered. No sentence is out of play, so the sentence "Bush is president of the US" is not out of play. According to Lewis's analysis of counterfactuals (1973), one of the following three counterfactuals is true in this context: "if the match were struck, Bush would be president of the US", "if the match were struck, Bush would not be president of the US", "if the match were struck, Bush could be president of the US". The analysis parts company with intuition and general opinion, which does not seem to accept any of the counterfactuals as true in that context.

The notion of a proposition being in play cannot be (easily) captured if one insists on working with a given fixed language. A natural move is to *explicitly* represent the language in play at a particular moment in the model of the logical structure effective at such moment (that is, the structure which allows evaluation of counterfactuals, fictional truths, belief changes, at that moment).[20] The resulting model of the logical structure involved in a given context is *local* in so far as it contains, and can only deal with, the sentences which are actually in play in that context.

---

[19] This model of context is defended in Stalnaker (1999).

[20] This sort of technique is employed, albeit in different forms, by Fagin and Halpern (1988), Modica and Rustichini (1999), and others working on the question of 'awareness' in the economics literature. The framework used here differs from the others in several important ways; see Hill (2008, 2010), for more details.

The construction of such a local model is relatively simple, and can, to a large extent, draw upon traditional models. Consider the example of counterfactuals and the analysis in terms of similarity between possible worlds (given this example, classical logic is assumed as the "underlying" logic). The local model of a context, where certain counterfactuals are in play is, more or less, as follows.[21] Take the set containing all and only the sentences in play in the context. Establish the appropriate logical relations between the sentences as they figure in the context: for example, the relationship "$p \leq q$" will be established if q is a logical consequence of p (according to the notion of consequence which applies in the context). As is well known (for classical logic), one obtains the structure of a Boolean algebra; indeed, for the sorts of case considered here, it will be an atomic Boolean algebra. The atoms of this algebra, or equivalently the maximal consistent sets of sentences, can be seen as possible worlds, or more precisely "small" possible worlds: *small* in so far as only the sentences in play in this context (and not all the sentences of some overarching language) have determinate truth values in these worlds. To allow evaluation of counterfactuals in this structure, traditional techniques can be employed: to the logical structure add a similarity relation on the small worlds, and evaluate the counterfactuals in the context in the ordinary way (the counterfactual is true if the consequent is true in all the closest small worlds where the antecedent is true).

Rather trivially, this local model yields an analysis of counterfactuals that does not suffer from the sort of problem raised above. In the context where the striking of the match is at issue, the question of the presidency of the US, and the sentence "Bush is the president of the US" are not in play. None of the problematic counterfactuals noted above can be formed in this context (they are not in play), and none of them are evaluated there as true or false.

The case of counterfactuals, and the use of a similarity relation between worlds is but an example. Once the appropriate *local* logical structure is set up, any machinery used in nonmonotonic logic or elsewhere (for example selection functions or epistemic entrenchment orders) can be added to analyse or account for the target phenomenon (counterfactuals, belief revision, fictional truth). Any logical theory which involves a supplementary structure on a Boolean or classical base can be *localised* in this way.

The advantage of these local models is not limited to the fact that they give a better account of the notion of being in play. On the contrary, it is their fruitfulness for dealing with the question of change that is most important. The identification of and restriction to the sentences in play at a given moment make no sense unless one admits that this set of sentences changes between moments and contexts, and one aims to take account of these changes. Indeed, the dynamics of the set of sentences in play has particularly attractive properties: as noted at the beginning of this section,

---

[21] A simplified exposition is given here; for a detailed, technical model, see Hill (2006, Ch. 5, 2008). Some of the specific issues, assumptions and arguments in defence of these assumptions, discussed in those papers, are left aside here.

it is apparently a prime example of something which changes *continuously*. Local models of the sort described above thus seem optimal tools for getting a preliminary grasp on the dynamics.

The basic intuition behind the continuity of the set of sentences in play is that, between consecutive moments or contexts, the sets of sentences in play overlap. However, using the sort of local models described above, more precise and supple relationships between the logical structures involved in consecutive moments can be defined. In particular, recall that the structure of the Boolean algebra for a given context captures the logical relationship between sentences *in so far as they appear in the context*; in other words, it captures their *local* logical structure. Stronger notions of continuity can be obtained by demanding not only that the sets of sentences in successive contexts overlap, but also that their logical structure does not change too drastically. For example, suppose at a point in his reading of a detective story, the reader comes to presuppose that, if the butler is guilty of the crime (A), the gardener is innocent (B). That this relation is a presupposition at that moment of the reading (in that context, so to speak) means that at all the (small) worlds which he can envisage in that context, and at which the butler is guilty, the gardener is innocent; *locally*, the guilt of the former *logically* implies the innocence of the latter (A implies B). A possible continuity constraint would state that such (local) logical relations cannot be suddenly reversed. In other words, in the successive context, or at the next moment, it cannot be that the reader presupposes that the butler's guilt implies the gardener's guilt (A implies not-B); that is, that at all the small worlds which he can envisage where the butler is guilty, the gardener is as well. The fact that such radical, abrupt changes of opinion seem unnatural speaks in favour of this as a sort of continuity which the dynamics of the logical structure respects. Note that it does not rule out the possibility that the reader can arrive in a context where he presupposes that the butler's guilt implies the gardener's, but only that the change cannot happen suddenly. And indeed, if one considers the way in which successive events in the detective story can lead the reader to accept that the guilt of the two is linked in this way (he discovers information which raises doubts as to the incompatibility of their respective guilt, and eventually, given further evidence and reflection, leads him to believe in a collaboration between the two), one quickly realises that one has to pass through a context where neither the implication from the butler's guilt to the gardener's innocence, nor the implication from the butler's guilt to the garden's guilt holds (when the incompatibility is put into doubt, the reader is in a context where there are small worlds where the butler is guilty and the gardener is not, and other worlds where there are both guilty: neither A implies B nor A implies not-B).

This is just one example in many. Given the local model as loosely described above, it is not too difficult to explicit a "calculus" of logical structures, where the different relationships between the logical structures with different sets of sentences in play are described. Well-known mathematical tools can be employed in this task. For example, recall that the local model of the context can be represented as an algebraic structure (indeed, a structure close to a Boolean algebra): one can thus tap into the armoury of concepts already developed for describing relationships between algebras, such as homomorphism, quotient and free product. These may be used to define different notions of proximity of logical structures and hence different notions of continuity. Indeed, the sorts of relationships and changes at work in the example above can be captured precisely and efficiently using only these concepts.[22]

The specificity of the local models described above is the locality of their language and the logical structure on it. However, any notion of continuity will equally have to take account of the specific element which is at issue: the relation of similarity between worlds, in the case of traditional theories of counterfactuals. The continuity principle requires that this extra factor is such that its change may be considered as continuous. A relation between local models which specifies a sense of proximity apt for an appropriate definition of continuity should not only take account of the sentences in play and the background logical structure (the set of small worlds), but also of the additional structure required to evaluate fictional truth, the changes of opinion upon reading a particular section of the novel, belief revision in general, counterfactuals, or whatever else is at issue. In this sense, the notion of similarity between worlds turns out once again to be rather clumsy, for the simple reason that such a relation is defined on a given set of (small) worlds, and is thus, at least at first glance, no longer defined when the set of (small) worlds changes. However, the set of (small) worlds will change as the context changes, and in particular as sentences fall out of play and new sentences come into play. For such reasons, it would be worth considering other factors, which could be more stable under changes in the set of sentences. To take but a simple example, if a notion of continuity like the one suggested above were to be used, the logical relationships among any set of sentences would be generally stable (that is, they would not change erratically) between contexts where all the sentences are in play: this more or less independently of the other sentences which are in play. Any factor which could be at least partially defined in terms of these logical relationships would thus be more likely to vary continuously between contexts. This would constitute a reason for developing an analysis of counterfactuals based on such a factor. This is simple to do: just hark back to the original analyses, proposed by the likes of Goodman (1954, Ch. 1) which consider counterfactuals to be strict implications with "relevant conditions" or "hidden premises" added to the antecedent. "If A were the case, then B would be the case" is analysed as "A and S implies B", where S is the conjunction of relevant conditions. These relevant conditions must be "independent" of the antecedent, in so far as (at least) "if A were the case, then S would not be the case" does *not* hold.[23] If this sense of independence is cashed out as *logical* independence, so that it is definable in terms of the logical relationships between sentences, then one obtains a factor which supports not only the evaluation of counterfactuals, but also a continuous dynamics. Evidently, a full theory must be more subtle than this simple sketch; for a first pass, see Hill (2006), Chs. 4 and 5.

---

[22] For a development of the discussion in the previous paragraph, and an introduction to the technical notions, see Hill (2006), §§5.1 and 5.2.

[23] This is what Goodman (1954), Ch. 1 calls "cotenability".

17.6 Conclusion

Fiction poses a set of problems relating to the notion of fictional truth and the reading of a fictional text. This links in with questions regarding counterfactuals and belief change. Questions which modern logic, in this instance nonmonotonic logic, would like, or should like, to deal with.

However, there are deep reasons why such logics have, so far, not had much to offer. In a word, logicians have generally assumed what needs to be understood, and so doing have evaded the difficult questions, to which responses are most urgently needed or desired. Accordingly, the interest and applicability of their work in other domains has been severely limited. The first aim of the paper is to establish this point.

This criticism should not be taken negatively, but as a call for a logic that does attempt to tackle the question of the exogenous factor which it assumes (similarity between worlds, epistemic entrenchment, selection function, to take but a few examples). This question is essentially that of the dynamics: understanding this factor amounts to giving an account of how it comes to receive a particular value in a particular context or at a particular moment. The second aim of the paper is to encourage the development of such theories (or logics).

To this end, some reflections on the project of a theory of dynamics have been offered. Such a theory should take as its first task that of defining a notion of continuity according to which changes in the logical machinery at work at successive moments is continuous.

This constraint weighs on the choice of the logical machinery employed in the analysis (in the example of counterfactuals and fictional truth, the similarity relation on possible worlds).

Suggestions for the sort of logical machinery which supports such a theory have been made, starting from the intuition that, if the sentences which are accepted or believed in a particular context or at a particular moment do not seem to change continuously, the set of sentences *in play* does seem to change continuously between successive contexts or moments. Furthermore, the local models one obtains by rendering explicit the sentences in play at a particular moment or in a particular context are relatively familiar mathematical structures: well-known operations can thus be employed to define a variety of relations between contexts, and subsequently a range of notions of continuity.

These last considerations are offered only as tentative suggestions. At most, they aim to show that the sort of theory called for here is possible, and that it can be developed employing techniques with which logicians are already familiar. The central moral of the paper is stated above: the project of understanding the dynamics is important and relevant in many domains. If other disciplines are to learn precision and rigour from logic, perhaps logic, in turn, could learn from them the worth of certain projects and aims.

References

Adams, E.W. 1975. *The logic of conditionals*. Dordrecht: D. Reidel Publishing Co.

Baltag, A., and L.S. Moss. 2004. Logic for epistemic programs. *Synthese* 60: 1–59.

Eco, U. 1985. *Lector in fabula*. Paris: Grasset & Fasquelle.

Edgington, D. 1995. On conditionals. *Mind* 104: 235–329.

Fagin, R., and J.Y. Halpern. 1988. Belief, awareness and limited reasoning. *Artificial Intelligence* 34: 39–76.

Gardenförs, P. 1988. *Knowledge in flux: Modeling the dynamics of epistemic states*. Cambridge: Bradford Books, MIT Press.

Goodman, N. 1954. *Fact, fiction and forecast*. Cambridge: Harvard University Press.

Goodman, N. 1970. Seven strictures on similarity. In *Experience and theory*, ed. L. Foster and J.W. Swanson. Amherst: University of Massachusetts Press.

Hill, B. 2006. Jouer avec le Faux. Recherches sur les processus mentaux à l'œuvre dans la lecture des textes de fiction. Doctorate Thesis, Université Paris I Panthéon-Sorbonne.

Hill, B. 2008. Towards a "sophisticated" model of belief dynamics. Part I: The General Framework. *Studia Logica* 89(1): 81–109.

Hill, B. 2010. Awareness dynamics. *Journal of Philosophical Logic* 39(2): 113–137.

Hintikka, J. 1962. *Knowledge and belief*. Ithaca: Cornell University Press.

Krasner, D.A. 2002. Semantics and fiction. *Erkenntnis* 57: 259–275.

Lewis, D.K. 1973a. *Counterfactuals*. Cambridge: Harvard University Press.

Lewis, D.K. 1973b. Causation. *The Journal of Philosophy* 70: 556–567.

Lewis, D.K. 1979. Counterfactual dependence and time's arrow. *Noûs* 13: 455–476.

Lewis, D.K. 1981. Ordering semantics and premise semantics for conditionals. *Journal of Philosophical Logic* 10(2): 217–234.

Lewis, D.K. 1983. Truth in fiction. In *Philosophical papers*, vol. I. Oxford: Oxford University Press.

Lowe, E.J. 1995. The truth about counterfactuals. *The Philosophical Quarterly* 45: 41–59.

Makinson, D. 2005. *Bridges from classical to nonmonotonic logic*. London: King's College University Press.

Makinson, D., and P. Gardenförs. 1991. Relations between the logic of theory change and nonmonotonic logic. In *The logic of theory change*, ed. A. Fuhrmann and M. Fuhrmann, 185–205. Berlin: Springer.

Matravers, D. 1995. Beliefs and fictional narrators. *Analysis* 55: 121–122.

Modica, S., and A. Rustichini. 1994. Awareness and partitional information structures. *Theory and Decision* 37: 107–124.

Modica, S., and A. Rustichini. 1999. Unawareness and partitional information structures. *Games and Economic Behaviour* 27: 265–298.

Pavel, T. 1981. Ontological issues in poetics: Speech acts and fictional worlds. *Journal of Aesthetics and Art Criticism* 40: 167–178.

Placek, T., and T. Müller. 2007. Counterfactuals and historical possibility. *Synthese* 154: 173–197.

Rott, H. 1999. Coherence and conservatism in the dynamics of belief. Part I: Finding the right framework. *Erkenntnis* 50: 387–412.

Rott, H. 2003. Coherence and conservatism in the dynamics of belief. Part II: Iterated belief change without dispositional coherence. *Journal of Logic and Computation* 13: 111–145.

Rott, H. 2006. Shifting priorities: Simple representations for 27 iterated theory change operators. In *Modality matters: Twenty-five essays in honour of Krister Segerberg*, Uppsala philosophical studies, vol. 53, ed. H. Lagerlund, S. Lindström, and R. Sliwinski, 359–384. Uppsala: Uppsala University Press.

Ryan, M.-L. 1991. Possible worlds and accessibility relations: A semantic typology of fiction. *Poetics Today* 12: 553–576.

Stalnaker, R.C. 1980. A defense of conditional excluded middle. In *Ifs: Conditionals, belief, decision, chance, and time*, ed. W.L. Harper, R. Stalnaker, and G. Pearce, 87–104. Dordrecht: D. Reidel Publishing Co.

Stalnaker, R.C. 1984. *Inquiry*. Cambridge: MIT Press.

Stalnaker, R.C. 1999. *Context and content: Essays on intentionality in speech and thought*. Oxford: Oxford University Press.

Sutherland, W.A. 1975. *Introduction to metric and topological spaces*. Oxford: OUP.

Van Benthem, J. 1996. *Exploring logical dynamics*. Stanford: CSLI Publications.

Woods, J. 1974. *The logic of fiction*. La Haye: Mouton.