UNIVERSITA' DEGLI STUDI DI PADOVA

Sede Amministrativa: Università degli Studi di Padova
Dipartimento di Filosofia

SCUOLA DI DOTTORATO DI RICERCA IN FILOSOFIA
INDIRIZZO: FILOSOFIA TEORETICA E PRATICA
CICLO: XX

REPRESENTING THE PAST

Direttore della Scuola: Ch.mo Prof. Enrico Berti
Supervisore: Ch.mo Prof. Antonio Da Re

Dottoranda: Dott. Ludovica Lorusso

TABLE OF CONTENTS

Chapter 3. Images of the past

INTRODUCTION

In the long history of humanities and science, human history and biological history have never been considered as a unique historical discipline. After centuries of separation between human sciences and natural sciences, recently there have been authors who have stressed the scientific character of historical inquiry in human history. But what about the humanistic character of biological reconstruction? Is it possible to find the same categories in the epistemology of historiography and in that one of phylogenetic systematics, the branch of biology that aims to reconstruct the historical relationships between species? Provided that it is possible to found historiography on scientific method, is it also possible that some aspects of the epistemology and methodology of human history can be adopted to understand the reconstruction of the phylogenetic past? My answer is Yes, and I shall give some arguments for this claim.

In my dissertation I define historical disciplines as disciplines that aim to give a historical interpretation of the evidence. Phylogenetic systematics is a historical discipline and therefore in my definition phylogenies should be thought of as historical interpretations of relationships between taxa.

I want to show why phylogenies should be considered as historical interpretations of the evidence, organized in a hierarchic and genealogical classification. The term ‘interpretation’ – which is considered to belong in non-objective knowledge – is not very welcome in science and in that philosophical realism that recognizes truth as the only epistemologically acceptable value. In phylogenetic systematics, to speak about the “true history” means in a strict sense to believe in the epistemological possibility to reconstruct the unique pattern that links the past to the present. To think about the knowledge of the biological past in this way does not make much sense, because of the fact that the evolutionary process has no memory, in the sense that the fact that evolution acts in a certain way in some organisms in some environments, it does not give us the knowledge of how, where and when evolution has been acting in the past. However, by using Oakeshott's thought about human history, the historian is not a memory, but a mind. Recalling does not represent a "better" knowledge than a conceptual organization of the evidence. Biologists do not need a memory of what happened, but a mind with the possibility of creating conceptual structures, thought experiments, ideal worlds of assumptions. And this is exactly what biologists do, they make assumptions inside what I shall name "ideal models". For example, they assume that evolution has been as a "constant and methodical creator", who has chosen to change things by following uniform intervals. By using this and other assumptions on the character of evolution, which constitute different ideal models, biologists can finally treat historical hypotheses with probability. Phylogenetic inference is founded on a probabilistic epistemology and this means, for instance, that predictions from the present to the past are probabilistic. Each historical hypothesis is characterized by a probability to be true. The present is projectible for the past only if we assume that some history is true, where by 'history' I mean the process generating the evidence. If a hypothesis about the future is probabilistically independent of any statement about the past and vice versa, if any hypothesis about the past is probabilistically independent of any statement about the present, to predict the past from the present and the present from the past is no possible.

Phylogenetic inference is based on a general hypothesis of *common ancestry*, which is also the hypothesis with the maximum cladistic parsimony, the one that maximizes the number of similarities between characters interpreted as homologies. Intuitively, the hypothesis of common ancestry is essential in every historical discipline, since without assuming this hypothesis to be the most plausible – which in a probabilistic epistemology means probable or likely – there would not be a history to reconstruct, but only a number of separate patterns independent of each others. Even separate patterns could constitute a genuine genealogical hypothesis, but the truth is that we like to think of history as a unique pattern or structure of events with only one origin. However, this hypothesis represents a sufficient explanation for the evolutionary evidence only inside evolutionary contexts of assumptions about branch transition probabilities. In other words, assumptions are needed to make plausible the hypothesis of common ancestry and therefore to make possible a biological history.

The assumptions or constraints on the evolutionary process constitute models used in probabilistic methods in order to give a probability to phylogenetic hypotheses.

For all these considerations, I suggest the need of a new epistemology, a sort of "illuminated" realism, in which conceptual frameworks are needed not only in the selection of evidence, but also in order to assign a probability to a phylogeny. Phylogenies are not naïve reconstructions of the past, but "representations" of possible histories, characterized by a probability when they are associated to ideal models on the evolutionary process.

Phylogenies are "images" of histories, with a meaning given by their models and characterized by a certain probability. Exactly like Weber's ideal types, each historical picture can not be found inside the reality itself, since it is just an *utopia*. It is not *the* historical reality, but it represents it in an ideal framework. Only inside the models reality becomes meaningful for us in a historical sense. The probabilistic way of representing history, characterized by its peculiar metaphysics, epistemology and method, represents a "mode" in which we can *abstract* the experience, where a "mode" is what Oakeshott calls a *form* of the experience, it is “the whole experience arrested at a certain point and at that point creating a homogeneous world of ideas”.

In phylogenetic inference, an important point is the choice of evidence, that is represented by matchings between special kinds of characters. I shall argue that while assumptions may be sufficient to entail the facts to be explained (the matchings), this provides only a surface explanation. In truth, the correlation of character states is the product of the a complex evolutionary history of species, moulded by the complex phenomena of adaptation to the environment, migration and isolation. In phylogenetic inference, the evidence for these phenomena should be studied and not mere matchings between character states. The interpretation of correlations could give a real epistemological value to phylogeny, expressed by a predictive success across different species and data sets in a historical perspective, which William Whewell has referred to more generally as the “consilience of inductions”. From the state of a character we could predict with a certain probability the state of other characters inside the species, in other species and between present and past species. So far phylogenetic systematics has been interested in constructing ideal models to make some histories sufficient to explain matchings. In the future, phylogenies should try to answer more complex and meaningful questions on the laws of evolution and the change of epistemology I suggest in my dissertation should be interpreted as an invitation to focus more on conceptual issues in this field, not just on empirical.

In chapter 1 I present and discuss the issues concerning the metaphysics and epistemology of biological classifications, like classification of species and classification of races, in particular human races. In standard methods of phylogenetic inference, the reconstruction of history begins with a classification into groups or clusters; in other words, a classification is *needed* to reconstruct a phylogeny. A biological classification can either have a historical meaning, or simply reflect the similarity between taxa, or reflect characteristics with an adaptive meaning. Some biological classifications and concepts have been traditionally founded on genealogical relationships between taxa and for this reason they are called “genealogical”, “phylogenetic” or “historical” concepts. For instance, the *cladistic* concept of species is one of the most used genealogical concept for this taxon. Even some concepts of human race are defined on a genealogical criterion, in the sense that two individuals from the same racial cluster are believed to share the same phylogenetic history.

Since most of methods in phylogenetic systematics are probabilistic, in this chapter I shall also introduce different probabilistic frameworks in which we justify our beliefs in hypotheses about the biological past.

In chapter 2 I present some epistemological, methodological and metaphysical issues in philosophy of human history. I shall use philosophy of human history for two main reasons. Firstly, in order to show that the dichotomy between a scientific and a historical thinking should change in a historical and non-historical thinking. Secondly, because the epistemology of human history is extremely useful to understand the epistemological issues surrounding phylogenetic inference. For instance, I shall use Max Weber’s ideal-types and Michael Oakeshott’s modes of experiences, inside a neo-Kantian epistemological perspective.

Chapter 3 shows that the dichotomy between history and science is misleading, by means of a analysis of the structure of phylogenetic inference. I show that phylogenetic systematics is a discipline based on a structure of ideal models and background assumptions on the character of evolution, in which phylogenetic hypotheses are characterized by a probability (more precisely, they give a probability to the evidence).

I first must acknowledge Giovanni Boniolo, who has made this dissertation possible and has always been encouraging my work during the last three years. I am extremely grateful to Elliott Sober, for his invaluable and generous assistance and to my friend Malcolm Forster, who has been so patient in teaching me anything I know on probability. I would like to thank Enrico Berti, Antonio Da Re, Luca Illetterati, Francesca Menegoni, the all staff of the Department of Philosophy in Padova for their precious help, and the University of Padova, for financial assistance.

And…Gabriella, my tireless mother, who has been supporting me in any way over the long haul.

CHAPTER 1

EVOLUTION AND CLASSIFICATION

> *From the first dawn of life, all organic beings are found to resemble each other in descending degrees, so that they can be classed in groups under groups. This classification is evidently not arbitrary like the grouping of the stars in constellations.*
>
> (Charles Darwin, *The Origin of Species*)

1. Classification-Taxonomy-Systematics

Before starting a discussion about phylogeny and questions related to phylogeny, let me clarify the meanings of some words.

First, phylogeny is a branch of biology connected with taxonomy and systematics. All of these three disciplines focus on different aspects of the problem of biological classification.

What is a biological classification? A biological classification or classification of living beings is often defined as *the activity that belongs exclusively to systematics*. Biologists who use the term ‘classification’ as the activity of systematics, consider the latter *the study of the diversity of things and of all relationships among them.* In accordance with this idea, systematics is an attempt to understand the evolutionary interrelationships of living things, trying to interpret the way in which life has diversified and changed over time. While classification is primarily the creation of names for groups, systematics goes beyond this to elucidate new theories of the mechanisms of evolution. The idea is that systematics is the field in which biologists try to reconstruct the evolutionary mechanisms and the patterns of events that have led to the distribution and diversity of life. Sometimes, those who consider systematics in this way divide it into these two branches: phylogeny and taxonomy. Phylogeny is the systematics in its large meaning of study of the biological past. Taxonomy, otherwise, is considered as a synonym of classification (see below). According with this subdivision, taxonomy is the pragmatic face of systematics, that is the branch of systematics with the only pragmatic aim of organizing living beings in groups.

However, systematics and biological classification are sometimes used with the same meaning. In this case they are both defined as *the activity of classifying organic beings in categories*. The activity of classifying can be considered in particular as belonging to the field of biology, but in general it is the activity of ordering objects inside any scientific and non-scientific methodology (Hennig, 1966). The approach to the real can be read as a classificatory approach. Observing objects is an activity of classification in the sense that it is an ordering activity reflecting the necessity to discover boundaries and relationships among things. "If each of the many things in the world were taken as distinct, unique, a thing in itself unrelated to any other thing, perception of the world would disintegrate into complete meaninglessness" (Simpson, 1961, p.2). Already in Aristotle there is the thought that the research of the first principles and essential definitions of any science does necessitate a classificatory process. In a Kantian framework, observation is just the first step of a process of conceptualization of the reality. In this sense, classification is the activity of organizing objects in the structure of our concepts and this structure is necessary to make our experience of the reality an observation as opposed to a mere perception. The understanding of reality is an active process of representing objects and connecting these representations toward progressive unification.

> Thought is knowledge by means of concepts. But concepts, as predicates of possible judgements, relate to some representation of a not yet determined object. Thus the concept of body means something, for example metal, which can be known by means of that concept. It is therefore a concept solely in virtue of its comprehending other representations, by means of which it can relate to objects. It is therefore, the predicate of a possible judgment, for instance, 'every metal is a body'. The functions of the understanding can, therefore, be discovered if we can give an exhaustive proposition of the functions of unity in judgements. (Kant, 1781-1787, p.106, B94)

Obviously we can or can not agree with the fact that our order and structure reflects the real order and structure of nature and we can either care or do not care about how nature is organized. In the knowledge of differences and discontinuities in nature, not every distinction that we like to consider real it is real. Categories in a classification can be considered as mere constructions or artifacts of our mind. In the first class of zoology at the National Museum of natural history in Paris, Lamarck said to his students to study biological classification without forgetting that all the divisions one needs are just artifacts and nature does not recognize any of them.

Scientists are usually looking for an agreement between classifications obtained by using different criteria, where this agreement is called *robustness* of the results. The robustness of the results is an analogous of the idea of the *overdetermination* of the measurements of observational quantities, which are independently "measured". For example, it is well known that our perception of visual depth is "overdetermined" by multiple cues, such as occlusion, disparity, binocular angle, looming and various monocular cues.

Unfortunately, what happens with classifications in biology is that different criteria often have different goals, that is they aim to describe different sides of the biological reality. Sometimes these criteria make sense in different epistemological frameworks. For example, the morphological and phenetic criteria[1] of species classification – which aim to classify organisms by using phenotypic characteristics – make sense in an instrumentalist framework. However, Mayr's biological criterion – which aim to classify organisms in an evolutionary perspective – makes sense in a realistic framework. For the first two criteria aim to classify species in a useful way, while the last one aims to find the true classification.

Finally, the term 'taxonomy' is used either with the same meaning of classification – as *the science of naming and classifying organisms* – or as *the theoretical study of classification, including its bases, principles, procedures, and rules* (Cfr. Simpson 1961). The term 'taxon' is usually defined as *a taxonomic group of any rank that is sufficiently distinct to be worthy of being assigned to a definite category*, where a *category* is defined as *a given rank or level in a hierarchic classification* (Mayr 1963, p.13).

1.1 The Species Concept

What is the problem in defining species? When Darwin proposed his theory of "Descent with modification", the idea of species in the *paradigm* of that period did not fit to his revolutionary framework. A species was considered a natural kind, linked to the idea of a Linnaean classification of immutable essences: in such species concept two

---

[1] Morphological and phenetic criteria are used in the homonymous species concepts. They both use phenotypic characters. While morphological criterion focuses on some key characters, the phenetic criterion uses all measurable characters.

individuals belong to the same species by virtue of their immutable essence, where the essence is made by the possessing of similar morphological properties. This theory of species is called *essentialism*.

Darwin thought differently about species, or better, he did not consider them too much. He was not interested in species, their properties, or in general in a classification of natural beings; he was interested in the mechanism of modification of species and the explanation of variation among species. His theory of evolution was a theory about this modification, about the *change* of natural things and not about their essences. In this new perspective, the essentialist species idea had become not necessary; moreover, that idea of species represented a problem for understanding the new theory of evolution. Interestingly, in 1860, in a letter to his friend C. Lyell, Darwin tried to convince him about the necessity of doubting the existence of species in order to sway the mind towards evolutionary theory.

After Darwin's theory, the problem of finding a non-essentialist definition of species arose. A species could no longer be considered as a category of organisms sharing a well defined immutable essence. The answer to this problem was "populational thinking". In populational thinking, species consist of populations. They have no essence, but "an internal genetic cohesion owing to the historically evolved genetic program that is shared by all members of the species" (Mayr, 1963, p.12): a species is a gene pool exposed in the time to the power of evolution. The populational thinking represents the theoretical basis of different species concepts, born in opposition to the old essentialist species idea. The most famous of these is the *biological species concept* (BSC), which sometimes (see Mayr) is considered as the sum of many different species definitions.

Let me briefly summarize the question of the species concept.

The *typological species concept* (called sometimes the *essentialist concept*) was the species concept used by Linnaeus and his followers (Cain, 1958). The idea behind this concept is that under the observed diversity of the real, there are a limited number of underlying universals, or types. Living beings are not considered in a dynamic perspective as different individuals capable of interacting with each other, but they are considered as different expressions of the same type, where this difference or variation is not a value but instead it represents the "imperfection" in the idea inside each species. The essence in each type is inferred from morphological similarity. Because morphological similarity is the criterion used in order to infer the essence of a species, the *morphological species concept* is considered by Mayr as a synonym of the typological species concept. However, I am going to show that they are not. For, while the typological concept assumes that the real is immutable (or at least in according with a largely accepted interpretation of it), the morphological concept does not make any assumption about the nature of world and can be considered only a pragmatic species concept in which the morphological criterion is used in a conventionalist way. Pragmatic in the sense that species categories have only to be useful and do not need to be true. Conventionalist (CC, see below) in the sense that the choice of one classification over another is just matter of arbitrary convention.

A kind of morphological species concept is what Lamarck and Darwin himself have in mind, when they speak about an "artificial construction invented because useful". In order to clarify his idea of the morphological species concept, Mayr makes a distinction between morphological species concept and morphological criterion. He refuses the first and accepts the second. As he says: "It must be emphasized that there is a complete difference between basing one's species concept on morphology and using morphological evidence as inference for the application of a biological species" (Mayr, 1969, p.25). The idea is that the morphological criterion can be rightly considered only a neutral tool, which can be used inside any species concept with a pragmatic purpose. In taxonomy phenotypic characteristics can be used – with other different kinds of evidence – in order to determine if a population deserves to be ranked as a species under the BSC. However, the morphological species concept uses the morphological criterion with the wrong purpose of inferring the essence of a species under the typological species concept.

The idea of the BSC is that a species consists of interacting populations characterized to have an internal “genetic cohesion” owing to the historically evolved genetic program that is shared by all members of the species. According to this concept, a species consists of populations and it has three properties: it is a *reproductive community*, in the sense that a species is created and sustained by barriers to the gene flow, and it is so possible to define it by reproductive isolation. From this first property of biological concept, a species amounts to a *genetic unit* consisting of a characteristic gene pool. Finally, a species is an *ecological unit*, in the sense that it is capable of interacting with other species with which it shares the environment. For some biologists, the last property represents a different species concept, called the *ecological* one, in which a species consists in populations sharing the same ecological niche. Interestingly, we can see that the ecological species concept for some aspects could be very close to the idea of species that Darwin had at his time. For, he had experience of species as populations separated because they had adapted to different environments through the pressure of natural selection.

Mayr calls the BSC the “theoretical” species concept, with the intention of stressing the distance from the morphological one - which for him cannot exist as a concept inside the theory of evolution. “The species definition that results from this theoretical species concept is: *Species are groups of interbreeding natural populations that are reproductively isolated from the other such groups*” (Mayr, 1963, p.12). Inside this concept species are populations included in a *temporal* project - the evolutionary project - and in a *spatial* context, which is their own ecological niche. The perspective of a spatial boundary to which species have become adapted has been interpreted in a different and more dynamic way with the theory of “niche construction”. The idea is that organisms shape their environments as well as being shaped by them. This idea comes out from two famous papers by Richard Lewontin from the early 1980s (Lewontin 1982, 1983) and it is now largely discussed in philosophy of biology. As Paul Griffiths writes:

> Niche construction is a challenge to what Lewontin called the ‘lock and key’ model of adaptation. In this metaphoric vision of evolution the adaptations of organisms are solutions (keys) to the problems posed by the environment (locks). Organisms are adapted to their ways of life because they were made to fit those ways of life. In place of this traditional metaphor of adaptation as ‘fit’, Lewontin suggested a metaphor of mutual construction. Organisms and their ecological niches are co-constructing and co-defining (Griffiths 2005, p.11).

By stressing the idea of interactions between organisms inside species and the idea of species as entities living in a spatio-temporal context, the new theory of “species as individuals” comes to the fore. In this theory species are not considered as a-temporal classes with members, nor simply as populations, but they are more considered as spatio-temporal entities like organisms. This new species concept has been proposed by Ghiselin and then accepted by other people like Hull. Today many biologists and philosophers hold it, considering species as a kind of *mereological sums* that are cohesive the way organisms are (about this topic see Brogaard, 2004). In defying species as a mereological sum they want to stress the possibilities of interaction in both directions inside a species. Not only from the species to a single population belonging to that species, but also from every single population to the species. Like in the case of niche construction, in this new perspective there is the purpose of discovering new properties inside the real, through the analysis of the biological complexity. Moreover, behind this concept there is the modern synthesis treatment of populations as ontological individuals, in which a particular logical relationship between organisms and populations was stressed. This relationship involving the parts with the whole was born in opposition of the idea of an individual membership in sets, classes, or kinds on the basis of shared intrinsic properties (see Gannett 2004). This idea of species-population as a mereological sum could be explained inside the framework of an integrative pluralism (see Mitchell, 2002), with two interacting levels: populations and the species.

The genetic aspect of the BSC has become an independent concept, the *genetic* species concept, which is based on the genetic similarity. The idea of species as a reproductive community has inspired the concept of *recognition*, focused on reproductive characteristics that preserve the gene pool integrity by avoiding the mating between species (see Paterson 1985, 1993).

It is worth pointing out that right now most of the species concepts are based on genetic similarity, ever if they have different purposes.

1.1.1 Genic and Genealogical View of Species

For many years Mayr's BSC has been the prevailing paradigm and, tied with this, the idea of allopatric speciation. Mayr argued (1942, 1963) that speciation occurs only in conditions of allopatry, that is in conditions of geographical separation. A consequence of this idea was that mechanisms of reproductive isolation (RI) were considered not as a primary target of natural selection and involved in speciation, but more as a by-products of allopatric divergence. In other words, RI was considered by Mayr as an effect resulting from other more primary divergence processes directly involved in speciation and its role was that one of keeping and not generating differentiation (speciation).

In the last years, RI has started to be considered as one of mechanisms directly involved in speciation (Coyne, 1992; Orr, 2001), and several studies have been focusing on discovering all mechanisms of adaptation evolving in sympatry, where sympatry is the condition in which species live in contact with each other. Moreover, in the last 15-20 years molecular research has shown that species experience some level of interbreeding and can share some regions of their genomes. In the new genetic framework, species and mechanisms of speciation and adaptation are not well defined inside the paradigm of BSC, and mechanisms of differentiation and adaptation must be reconsidered outside the rigid criterion of RI, as features that make species more than merely divergent populations of the same species. However, the new *genic view of species* and *speciation* does not offer much more room for these new mechanisms, since it is conceptually and methodologically similar to the old BSC. This is one of the genic definitions of species: 'Species are groups that are differentially adapted and, upon contact, are not able to share genes controlling these adaptive characters, by direct exchange or through intermediate hybrid populations' (Wu, 2001). We can see how the idea of RI is still central in this concept of species. Species are evolved and so definable as differentiated groups of populations if and only if in conditions of sympatry adaptive differences between them can prevent their fusion. The idea of reproductive boundaries between species seems to be a necessary condition also in the genic definition of them. It follows that contact is a necessary condition in order to recognize two populations as two different species. Note that in this definition adaptation is considered - like in Mayr's BSC – in its role of keeping species separated from an ecological point of view. Since species are characterized by genes for different adaptive characters, they will not be able to share the same ecological niche. At the same time, geographical separation seems to be a necessary condition for speciation in Wu's genic definition of species, despite recent advances suggesting that non-allopatric speciation is also possible and perhaps even common (Shaw, 2001). Summarizing, there are three characteristics of this view: (1) adaptation is the necessary force of generating speciation and keeping species separated in conditions of sympatry; (2) genes of adaptation are the only ones involved in mechanisms of speciation and determination of species; (3) species are identifiable only in a sympatric context.

Baum and Shaw (1995) suggested a *genealogical view of species* and *speciation*, in which a non-adaptive concept of species is considered, in the sense that adaptive features of species are not a necessary condition of their existence. A genealogical species is "a basal group of organisms, whose members share exclusive genealogical relationship. Genealogical relationship in this context refers to how far back in time members of a group must trace their ancestry (e.g. assessed through composite gene-genealogies) before they are connected by progenitors in common". In this species concept any single evolutionary force (selection, drift), gene, and geographical context is considered by itself sufficient but not necessary to give rise to a process of speciation and to keep species separated. The table below shows the major species concepts proposed in the literature and the biological factors considered in their definition.

Species concepts	Focus
Biological	Reproductive and adaptive characteristics, focus on any kind of similarity relevant to adaptation and reproduction
Cladistic	Phylogenetic relationships
Cohesion	Reproductive and ecological characteristics; genetic similarity
Ecological	Ecological characteristics
Evolutionary	Ecological, adaptive characteristics and phylogenetic relationships
Genetic	Genetic similarity
Morphological	Genetic similarity as manifested in phenotypic similarity; focus on key characters
Phenetic	Genetic similarity as manifested in phenotypic similarity; all measurable characters considered
Phylogenetic	Phylogenetic relationships
Recognition	Reproductive characteristics and genetic similarity; focus on behavioural traits
Genic	Genes controlling adaptive characters
Genealogical	Genealogical relationships

Table 1. A list of some of the major species concepts proposed in the literature and the biological factors considered in their definition (Pigliucci and Kaplan 2006, modified)

1.2 Realism and Nominalism about Species

Living things are classified into distinct units, called species: how can this practice be justified? Do species really exist as units in nature? Or they are merely divisions invented by taxonomists? (Ridley, 1985, p.89).

The question about the reality of species has been faced in the history of biology under the assumption that there exists *one* concept describing the real biological classification. In particular, two species concepts have been compared: the typological and the populational concept. For Mayr, only species that are described inside the BSC are real. Sometimes he makes a distinction among the BSC, the typological – which he also calls morphological – and the *nominalistic species concept* (Mayr, 1963). The idea in this concept is that species are nothing else but “names”. We have seen that he considers the morphological species concept as a concept that - as the typological one - uses the morphological criterion inside the essentialist theory. However, in other situations he associates the typological (morphological) concept with the nominalistic one, arguing that in both of them species would merely be arbitrary mental constructs. Therefore, for him, typological species do exist only like names that biologists give to them and they do not have an independent existence in reality. For, the use of an arbitrary criterion in grouping classes is strongly linked with the idea that species are not something different from other classes of objects, or, in other words, that they are not different from natural kinds:

> The nominalists deny the existence of real universal. For them only individuals exist; species are man-made abstractions. (…) Any naturalist, whether a primitive native or a trained population geneticist, knows that this is simply not true. *Species of animals are not human constructs*, nor are they types in the sense of Plato and Aristotle; but they are something for which there is no equivalent in the realm inanimate objects (Mayr, 1963, p.12. Italics added).

This is an argument *ad populum* and it is the same kind of argument used by Simpson (see below) for justifying the hierarchic classification of organisms. His argument is that species do exist because any naturalist knows that species do exist. Maybe there is knowledge to which only taxonomists and naturalists have access. This idea could make sense inside a naturalized epistemology for naturalists! A better argument could be that, because naturalists can manipulate species creating unfertile individuals from interbreeding, then species must exist. Another argument is related with the idea of robustness of classifications (see above): species exist since naturalists do find the same species using different criteria of classification. This is the idea that Hacking expresses in his article "Do we see through a microscope?" (1981). He argues that because one is able to find the same results by using different kind of scientific tools as microscopes, then one should suppose that behind this robustness of the results there is something "real". "Real" is what we are able *to manipulate* and to find by using *independent* methods. Unfortunately, this reasoning does not work for species. Since the point is that different criteria of classification actually lead us to different species! So, why should we believe that the only correct criterion is the biological one and that using this criterion we can have access to the real species? Mayr does not answer this question as Simpson does not find a justification for using the hierarchic criterion of classification (see below). They both think that an agreement in the beliefs of naturalists is sufficient to know the biological world.

In another paper, Mayr stresses again the idea that with his BSC the species as population finds its *reality* not in its essence – which is not real! – but in the individuals of which it is composed and in their variation. He expresses this point with the following words:

> […] there are a limited number of fixed, unchangeable "ideas" underlying the observed variability (in nature), with the *eidos* (idea) being the only thing that is fixed and real, while the observed variability has no more reality than the shadows of an object on the cave wall […In contrast], the populationist stresses the uniqueness of everything in the organic world […] All organisms and organic phenomena are composed of unique features and can be described collectively only in statistical terms. Individuals, or any kind of organic entities, form populations of which we can determine the arithmetic mean and the statistics of variation. Averages are merely statistical abstractions, only the individuals of which the population are composed have reality. The ultimate conclusions of the population thinker and of the typologist are precisely the opposite. For the typologist the type is real and the variation an illusion, while

> for the populationist, the type (average) is an abstraction and only the variation is real. No two ways of looking at nature could be more different (Mayr, 1969, pp.28-9).

Here Mayr does not compare the BSC with the nominalistic view of species anymore. Instead he considers the difference between a typological and a populational way of thinking. A species is at the same time a population and the individuals in that population. While the *type* is the statistical mean of the population, and so it is an abstract man-made immutable object, the individuals of the populations are real. Are the ideas of type and individual incompatible in defining a species? No. They are just different aspects of a species. We can think about species in two different ways, depending on what we are looking for. If we are interested in knowing the mean of the population of a species, we must think typologically and use statistics. If we are interested in understanding the evolutionary aspects of the species, we must think about the individuals and their variation.

If we are interested in populations we should choose an instrumentalist framework. Instead, if we are interested in individuals, we should choose a realist one.

After these considerations, I suggest that from being a typological (essentialist, morphological) thinker of species does not follow to have a nominalistic position about species, because thinking about species as types does not imply thinking about them as artificial man-made entities. The type is only an *idealization* as the *mean* of characteristics of individuals of a population. Using idealizations in science is perfectly compatible with a non-nominalistic picture of reality.

1.3 Conventionalism in Species

Another position on species is conventionalism. There is a kind of conventionalism that can be called *common view conventionalism* (CC), in which the choice of concepts, principles, hypotheses is only a matter of convention, in the strong sense that using different concepts in different situations is totally *arbitrary*. In what sense arbitrary? Arbitrary in respect of the knowledge of the world. We choose a theory on the basis of usefulness, beauty, psychology, etc., not considering the possibility that our choice of it must be influenced by our understanding of the world. In other words, this kind of conventionalism could have its explication inside a pragmatic, aesthetical, psychological framework.

Inside this kind of conventionalism the species problem dissolves, since the choice of a species concept as opposed as another is only a matter of arbitrary convention and it is totally independent of any kind of epistemic or ontological consideration.

However, I want to suggest here that the Poincaré's conventionalism (PC) is a better framework in which to understand the problem of the species concept.

In the J-H. Poincaré's conventionalism (PC), the idea of conventionalism does not entail arbitrariness. For the philosopher in his work *La science et l'hypothèse* argues that the freedom in the choice of different theories and principles is never arbitrary, because if it were, science would be maybe certain but without any information and value. His idea of conventionalism is based on the thought that in science there are no true theories *a priori* in the Kantian sense of "logically necessary". At the same time the choice of a theory does not depend on how the world really is, but it is only the more convenient choice we make in our understanding of the world. "Conventional" in PC does not mean usefulness without knowledge. From his idea of considering principles as conventions it does not follow that the choice of principles is arbitrary, since the experiences determine understanding and understanding leads you to choose one principle instead of another. The "convention" reflects a kind of compromise between our understanding and experience. The only aspect in common between the CC and PC is the idea that the aim of science is not to reach a unique representation. In PC, the choice cannot be made a priori and that the choice of a theory over another is made on an epistemological level and not ontological. For the world cannot tell us what is the “truest theory”, but only what is the most convenient for us in order to understand it.

He imagined a different and fantastic world where people could have impressions totally different from ours and so they would be guided by the experience to create another kind of geometry. But, how he stresses, languages from different geometries are the same, so that they are perfectly compatible.

> Experiment guides us in this choice, which it does not impose on us. It tells us not what is the truest, but what is the most convenient geometry. It will be noticed that my description of these fantastic worlds has required no language other than that of ordinary geometry. Then, were we transported to those worlds, there would be no need to change that language. Beings educated there would no doubt find it more convenient to create a geometry different from ours, and better adapted to their impressions; but as for us, in the presence of the same impressions, it is certain that we should not find it more convenient to make a change. (Poincaré 1902, pp.70-71).

The idea behind this kind of conventionalism is close to the contemporary idea of compatible pluralism (see below), in the sense that in both there is no arbitrariness in choosing some theories instead of others. Moreover, there is the idea of compatibility between different theories.

Different biological classifications are only different representations of the biological world. The choice of a species concept is not a choice of a specific ontology, but just a choice of a convenient representation. However, this does not mean that it must be arbitrary, since we saw that it depends on what we are looking for in different situations.

1.4 Choosing Revolution

As I showed above, the common view in the history of biology is based on the idea that Darwinian “revolution” has made it possible to conceive a new kind of species concept and a new kind of classification: the BSC and the genealogical classification. The idea of revolution has been taken as implying an incompatibility and incommensurability between the pre-Darwinian essentialistic and the post-Darwinian populational view of reality, in all possible aspects. The BSC and the genealogical classification make sense inside the Darwinian revolutionary theory about a reality, consisting of entities unique in their diversity and considered for the first time in their historical dimension, as entities changing in the time. The idea is that there is an old framework of fixity versus a new dynamic one, where there is not room for classes of immutable natural kinds, but only for mereological sums (see above) of populations changing in the time.

However, in some relatively old papers and in more recent ones we can find different views of the relation between essentialism and post-essentialism that challenge the common tendency of considering them incompatible under any possible aspect.

There are some philosophers who do not accept the separation of the history of species into a pre-Darwinian period of essentialism and fixity and a post-Darwinian period of populational thinking and change. For example, Sober (1980) analyzes population thinking and the idea of essentialism, holding that essentialism is not a conception necessarily linked with the idea of fixity of nature. He proposes an interpretation of essentialism as an alternative and scientifically respectable way of understanding the existence of variation in nature.

Also Winsor (2003) proposes an alternative view. She argues that typology and the typological method in pre-Darwinian taxonomy should not be associated with essentialism. Her argument is that essentialism has been related to systematics in the pre-Darwinian era because of the failure to separate ontology and epistemology: the point is that one cannot assume that naturalists in the pre-Darwinian era were essentialists ontologically (in their world view) and not just essentialists epistemologically (in their method). For, she tries to show that there is a continuum in the taxonomic methods and that the method of types can be considered today as a method of clustering, without any ontological assumptions holding the link between the typological method and essentialism. For other philosophers (Stamos, 2005) pre-Darwinian taxonomists can be considered clusterists and not necessarily essentialists about what concerns the higher taxa (genera, families, orders, etc.), but not the species category. Stamos argues that naturalists and taxonomists in the pre-Darwinian era were ontologically nominalists and epistemologically clusterists about the higher taxa, but they were realists about the species. For him, to be realist does not necessarily entail that one is an essentialist, clusterist, or whatever else: "to be realist about species one needs only to believe that species are extra-mentally real. Species might be essentialist classes, cluster classes, groups, individuals, or whatever else, and they might evolve into other species or not" (Stamos, 2005, p.84). However, fixism implies realism: "To say that a species is fixed, however, is to narrow the range of species realism" (Stamos, 2005, p.84). Finally, for Stamos, essentialism entails fixism, at least about species, and fixism entails realism.

These new reconstructions of the history of biology show new kinds of correlations among biological concepts. In this new way of seeing the history of biology, theories considered incompatible, if differently interpreted – for example on different levels of knowledge (methodological, epistemological, ontological point of view) – become compatible and can interact with each other. The idea is to use pluralism in the history and philosophy of science. According to the pluralistic view of science, the diversity of explanations, representations, models and theoretical approaches is no longer considered correlated with the immaturity of science (Kuhn, 1962), but with the complexity of the subject matter. In this sense, the strategy and goal of attaining scientific knowledge does not consist in one unified explanation, theory or model. The structure of scientific knowledge must be founded on many different alternative explanations, theories and models and on different levels inside them.

The idea of maintaining the competition among different theories in order to hasten the progress of science, is not a new idea, since we can find it in the work of Feyerabend (1981) and Lakatos (1970). However, this "new" kind of pluralism is quite far from the pluralism of Feyerabend and Lakatos. The main idea here is not to maintain the competition among different research programs or theories with the ultimate aim of resolving differences and conflicts among alternatives theories. The new perspective looks for possible links in the network of different theories or research programs in science and focus on the communication among different points of view even those that are very distant from each other. This new kind of pluralism is called compatible pluralism (see Mitchell, 2002).

> The slogan here [in the competitive pluralism] might be 'Pluralism: the way to Unite' These accounts of competitive pluralism presume that pluralism is temporary and strategic, but ultimately eliminable. While this analysis correctly describes some of the diversity of models and explanations found in contemporary science, it fails to capture all of it. The remainder is constituted by compatible, not mutually exclusive, alternatives (Mitchell, 2002, pp.56-57).

1.5 A *Non-Arbitrary* Classification of Organic beings: the *Cladistic Concept of Species*

The distinction between reality and abstraction in a species concept is related with the distinction between natural and artificial or non-arbitrary and arbitrary classification of species in the higher taxa. Species as a historical entity and as a population (the two main characteristics of the populational thinking) is evidently linked with the necessity of a new concept of classification, no longer based on sharing morphological properties, but based on the history of the species: a genealogical and not morphological classification. Three species concepts are based on the idea of a genealogical classification: the *cladistic*, the *evolutionary* and the *phylogenetic* species concept.

Nevertheless, the morphological criterion is still the most used in systematics.

This is not surprising or something to be afraid of. We have seen before that Mayr himself holds that it is coherent to use morphological criteria in order to individuate the species inside the BSC. The idea is that one can use morphological criteria in individuating species or in making a higher classification because it does not entail essentialist. A species will always be different from a natural kind because it can change. While the nature of gold will be always the same, a species can evolve in something new, changing its properties and so its nature. So, a species cannot be a natural kind like a chemical element. This is one of the most famous arguments used against essentialism. Unfortunately, we know that also for chemical elements changes (transmutations) are possible. So, the idea of changing does not support the ontological distance of a species from a chemical element and in general from a natural kind.

A gradual evolution is also considered to be an argument against essentialism.

A species can evolve following two different models. The first one is the *anagenesis* (Figure 1b), according to which a species *gradually* evolves into a different species. In anagenesis the idea is that one species can become another species without splitting into two new species. The other model of evolution is the *cladogenesis* (Figure 1a) This model describes a branching process (*κλαδοζ*=branch), in which one species splits into two or more different daughter species.

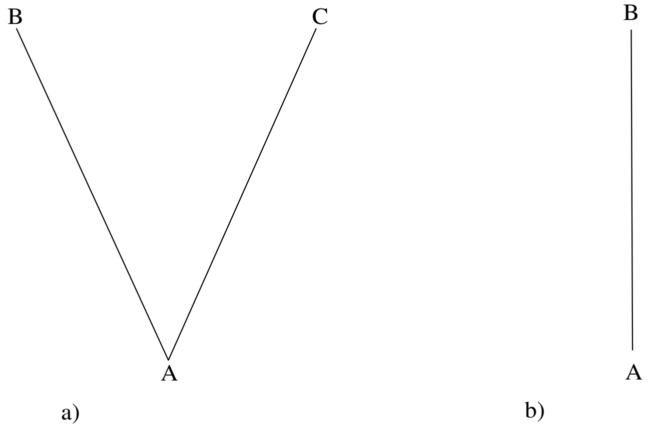


Figure 1. Cladogenesis & Anagenesis

The historical classification of species, that for Darwin was represented by the *tree of life,* requires that the number of species increases. This requirement can be satisfied by both the processes of anagenesis and cladogenesis. For anagenesis, however, an additional assumption is required. We have to assume that the ancestral species does not disappear. The process of anagenesis can be understood by thinking about the process of *gene duplication* inside genes. In our genome there are several populations of the same gene. At one point some of these genes can develop a different function, mostly depending on the site at which they are located inside the genome. The other genes having the old function still exist inside the genome. In the same way, a species is a set of populations, where it is always possible that one of these populations, under the pressure of the environment, differentiates to the point of becoming another species. This idea can be easily understood if we focus on the ecological species concept, where a population B evolves into another species A when they become completely adapted to two different ecological niches. Here the idea of speciation is linked to the concept of ecological niche and adaptation to the environmental and can be seen as a process in which a divergent selection pushes a population inside a species to diverge and eventually become another species. Therefore, if we assume that the ancestral species does not disappear, then also in anagenesis we have that the number of species increases. See Figure 2.

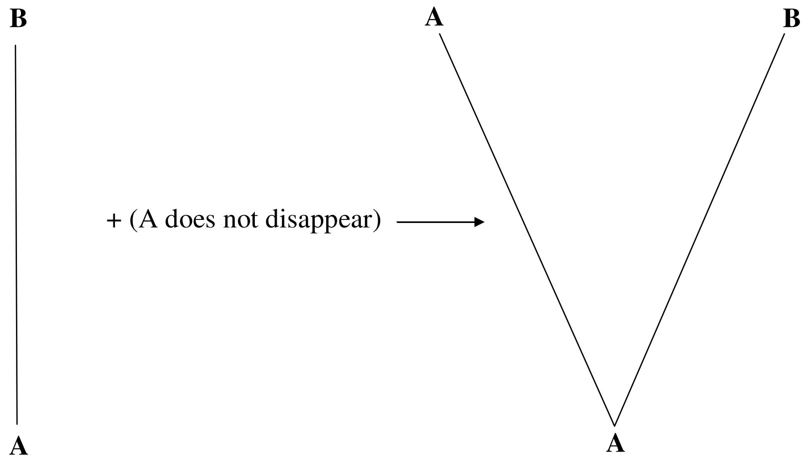


Figure 2

Another point about anagenesis is that by considering evolution by anagenesis to trace a non-arbitrary boundary among different species is impossible (see e.g. Sober, 2000). On the contrary, by considering cladogenesis we can individuate species in a non-arbitrary way. Above, I presented a definition of anagenesis, in which the idea of a gradual evolution is stressed. However, in Darwin the idea of a gradual evolutionary process and the idea of a genealogical classification of species, represented by a tree, were perfectly compatible. We can make two observations. First, to say that the idea of a gradual evolution represents a practical problem in taxonomy in individuating species and higher taxa is false because all classifications – since they are representations of evolution – can be thought of as gradual. The idea of gradual evolution is compatible with both a branching and a non-branching pattern, since I can consider the changing inside any branch of a tree to be gradual. I can think about the process inside any branch as acting in different ways. To be gradual all the time along the branch; to be gradual and at some points to become very rapid and to cause the species to split in another quickly. I can also think about a process of stasis in which at some point a new species arises quickly. The last one is an alternative version of the idea of the *punctuated equilibrium*. In this model of evolution the life of a species consists in a relatively sudden appearance followed by a long period of stasis terminated either by extinction or by splitting into daughter species. In according with this model of evolution, the typical life story of a species involves defined origins and terminations. The case I am considering is different because at the end of the process of stasis we have a new species but the old one does not disappear.

Second, the arbitrariness of a classification does not depend on the kind of evolutionary model we consider, but only on the justification we are able to give to it. Third, classification is about tracing boundaries between organisms and any line we can draw in order to individuate this boundary can be arbitrary, independently of the nature of the process we are assuming.

2. Classification of Races

In this section I want to analyze the classification of races in humans. So far this issue has been often analysed in the fields of social and moral philosophy. However, here, I am going to analyse the concept of race in a new perspective, that is from an epistemological point of view. Like we have seen in classification of species, also for racial classifications there is the ontological problem of considering race either as a real entity or as a human construction. I shall briefly present the epistemological issues concerning the concept of race.

In medicine and pharmacotherapy racial categories are frequently offered as the best explanation for differences in risk of diseases and drug responsiveness. However, an epistemological discussion about the role of racial categories in scientific reasoning is rare in these fields. I shall analyse the epistemological status of the concept of racial categories. The question I aim to answer is whether and when races belong in biological explanation. I define race as a genetic discontinuity between human categories. I shall show that differences in risk of diseases among racial categories are consistent with different explanations, only one of which implies the existence of race. Moreover, I shall show that race is not a projectible concept. 'Projectibility' is a term used by Goodman (1954) for a property of an object to make successful projections. In conclusion, I shall argue that the concept of race is not justified and it is not projectible, even though scientists use it in their explanations.

2.1 The Concept of Human Cluster

Nowadays there is a wide discussion surrounding the concept of cluster in human biology and in particular the concept of racial cluster. Until now the most common question about race has been: Do races exist? This is an ontological problem, which for many people should be answered inside science, since the race concept is a scientific concept. For others, however, it has both a cultural and scientific dimension and the answer should involve philosophy as well. The answer coming from science is controversial. One of the main reasons for this disagreement among scientists is the semantic ambiguity of the term. The term 'race' is used with many different meanings and it is difficult to reach an agreement for the existence of something that means something different for different people. Therefore, there is another question that should come before the question about the existence of race. This question is: How do we define race? Finally, there is the issue I am going to answer here, which is not very well discussed in the scientific and philosophical literature. This is the epistemological issue, concerning the justification of the concept of race in biological explanation and its projectibility. One could express it with the following questions: Is the concept of race introduced into biological explanation in a justifiable way? Is it projectible?

I have presented three kinds of issues concerning race: the ontological, semantic, and epistemological one. All these problems acquire a particular significance in relation to a fourth question: Should the concept of race be used in biology? For all the three issues about race have become central in philosophy and science because the use of the race concept among humans is historically linked to ethical and social issues.

In philosophy almost every discussion about the use of race in biology is focused on the ontological issue. Roughly the most popular argument is: races do not exist, therefore the concept of race must not be used in biology. People who deny the existence of races can hold two different positions. In the first ontological position against races, races are considered social constructs. In the second one, instead, they represent the product of both social and biological properties. While in the first position any biological value of race is denied, the second one denies the possibility of a separation of social and biological properties in the definition of race. On the other hand, other philosophers reify races, by searching for viable conceptions of them[2]. By reification I mean the creation of a race concept with a new ontological status.

Unfortunately, the ontological argument does not work very well in eliminating the use of race concept from biology. The reason for its failure is due to the ambiguous way the concept is used in science. In biology scientists who use races hold an instrumentalist position by claiming that races should be used because they are useful, independently from their existence. Apparently they avoid the ontological problem,

---

[2] About the first position see e.g. Goldberg 1993. About the second position see e.g. Gannett 2004 and Hacking 2006. About a possible reification of the race concept see e.g. Andreasen 2000.

since they do not make a claim about the existence of race explicitly. What they say is only that a racial classification of humans is useful, therefore it should be used. But we shall see that their neutrality about the existence of races is not genuine and it can survive only inside a semantic ambiguity and the absence of an epistemological discussion under which they operate. A clarification of the semantic issue and the epistemological role of races in biological explanation is needed for a justified repudiation of the racial concept in the scientific reasoning.

Before starting my analysis, I need to make an important distinction, between a race and a racial cluster. I have just said that scientists like to claim that they use race in an instrumentalist way, without an explicit ontological commitment. When we find in scientific papers the word ‘race’, this term has not a declared ontological meaning and it can just refer to the operational concept of racial cluster, that is roughly a concept built on people’s self-declarations on their own ethnicity. However, I want to refer to the term ‘race’ with a precise ontological meaning, which characterizes the genetic definition of race (see §2.2). While any race is a racial cluster, not any racial cluster is a race. For I shall claim that a difference in genotypes among human clusters should be considered the necessary and sufficient condition for a racial cluster to be a race.

What is a racial cluster and in general a human cluster? A human cluster is simply defined as a class of humans sharing properties. Among the properties involved in clustering there is a “main property”, which is usually a phenotypic, cultural, or geographical property. This property allows us to choose the first boundary cluster and it is fundamental for the research because it is specified by the question being asked.

> Note, however, that it is extremely informative because in this way the geneticist-observer has already chosen a first kind of boundary, and a first cluster on which to perform the investigation. This means that geneticists usually start their work from what we might call the Main Cluster (MC), which is related to their research

> questions and which is chosen on the ground of phenotypic, cultural, religious, … properties, that might be called the main properties (Lorusso and Boniolo, forthcoming).

A racial cluster is doubtless the most used human cluster in current human biology. Phenotypic and geographical properties are mostly used to make it. While phenotypic properties are properties related to the colour of the skin, the length of bones, like for example “to be black”, geographical properties refer to the origin from a specific region, like for example “to be African”. Racial clusters are constructed by using a combination of properties, usually correlated, like the colour of the skin and the geographical origin (see for example the African-American population in table 2).

Even genetic properties can be involved in clustering humans, with the possibility of creating several genetic groups by choosing different genetic systems (nuclear DNA, mitochondrial DNA), genetic markers (SNPs, microsatellites), and sets of traits inside each genetic marker. However, even when we use genetic properties, these properties are investigated inside a main cluster which has already been fixed on the basis of non-genetic properties[3].

Whatever properties can be used in clustering humans, a cluster in itself is not problematic from either a methodological or an ethical point of view, but the problem is in the meaning we give to it and how we use it.

---

[3] See about methodological issues in the human clusters construction, Lorusso & Boniolo, forthcoming.

Population	Property
Mediterranean population	Geographical
Ashkenazi population	Religious
African-American population	Phenotypic & Geographical

Table 2. Clusters on the basis of some main properties

In biomedical research like in epidemiology, pharmacogenetics, and forensic science, racial clusters are widely used with the justification that they have a heuristic purpose. In particular, in biomedical research, they are involved in the explanation for differences in disease rates.

Some scientists argue that the risk of social and cultural discrimination and that one of undervaluing the diversity among individuals within groups should be weighed against the fact that “in epidemiologic and clinical research racial and ethnic categories are useful for generating and exploring hypotheses about environmental and genetic risk factors, as well as interactions between risk factors, for important medical outcomes exploring hypotheses” (Burchard et al., 2003).

I am going to question the usefulness of racial classification in biology, and in particular the explanatory and projectible value of races in biology. Is racial classification needed in biology? Can it be projectible, in the sense of being useful for generalization and prediction? Before facing the epistemological issue about the justification of race in biomedical research, I shall roughly present the problem concerning the definition of race and the problem concerning the existence of race, respectively the semantic and ontological problem.

2.2 The Semantic Issue: How to Define Race

The first issue about race I am going to analyse is the semantic one, which is the problem of the meaning of the term ‘race’. There are two main kinds of definitions:

1. An instrumentalist definition, which considers non-genetic properties as heuristically sufficient conditions for defining races.
2. A realist genetic definition, which uses genetic properties to define races. Genetic properties can be either differences in frequencies or differences in the form of discontinuities.

In definition 1 races are defined by non-genetic properties like geographical, phenotypic, cultural ones. Who holds this definition is not concerned about the existence of human groups, but only about possible uses of racial classifications, in particular their explanatory power and projectibility for risk of diseases.

The definition 2 involves shared genetic characteristics and it has metaphysical implications.

A difference in the form of continuity consists in a difference in frequency of genotypes among populations, in the sense that the same genotypes are found in different populations with different frequency.

A difference in the form of discontinuity consists in a difference in genotypes among populations, in the sense that the same genotypes are never found in different populations. This is an idealized case, since in the reality we could only have the situation in which the same genotypes are rarely found in different populations, with frequencies close to 1 in one population and 0 in another. By genotype I mean a specific number (1, 2, …, n) of genetic traits inside a genome.

Consider two populations defined by non-genetic properties. For you need to fix a main cluster first, defined for example by geographical or phenotypic properties and then you consider the genetic properties of those populations.

There are two kinds of discontinuity that I am going to consider in my analysis, a discontinuity for a single trait and a discontinuity in the whole genome:

- We have a discontinuity for a single trait if, given the two populations, that polymorphic[4] trait in an individual's genome is found exclusively in one of the populations. Therefore, from this trait it is possible to unambiguously predict the population to which the individual belongs.
- We have a discontinuity in the whole genome if, given the two populations, the complete set of all polymorphic traits in an individual's genome is found exclusively in one of the populations and from this set it is possible to unambiguously predict the population to which the individual belongs.

As it will be explained in the next section, a discontinuity of one single trait implies a discontinuity in the whole genome. Therefore in both cases of discontinuity, an unambiguous human classification is possible.

I shall adopt a realist genetic definition of race, which assumes the existence of discontinuity in genetic properties among racial clusters.

2.3 The Ontological Issue

> Different genetic polymorphisms are differently distributed over the planet, but their distributions are not generally correlated. This is a consequence of the fact that gene flow, rather than isolation, is the main evolutionary force shaping human

---

[4] A trait in the genome is defined polymorphic if occurs in different forms in a population with a frequency higher than 1%

> genome diversity, suggesting, therefore, a genetic landscape dominated by a continuous change among populations. As a consequence, clear-cut genetic boundaries between human clusters cannot be unambiguously identified. Only the discontinuity among human individuals can be unambiguously identified (Barbujani and Belle, 2006. Italics added)

In order to understand the distinction between continuity and discontinuity and between discontinuity for a single trait and whole genome discontinuity, let us consider two simple examples. In the first one we have a polymorphic trait G, which can be in two forms 1 and 2 and these forms are differently distributed among the two populations 1 and 2. Consider two different cases for its distribution:

	G1	G2
P1	1/2	1/2
P2	3/4	1/4

Continuity

	G1	G2
P1	1	0
P2	0	1

Discontinuity

Figure 3. One Trait Continuity and Discontinuity

Yet, let us make the situation more complex and realistic by adding another trait H, which can be in the two forms 1 and 2 inside the whole population P (P1+P2). In this case the two sub-populations P1 and P2 would represent two unambiguous groups of the population P for the trait G. While G is discontinuous, H is continuous between the two sub-populations. By extrapolating this case with two traits for all polymorphic traits in a genome, we shall have a discontinuity in the whole genome variation.

	(G1,H1)	(G1,H2)	(G2,H1)	(G2,H2)
P1	1/2	1/2	0	0
P2	0	0	1/2	1/2

Figure 4. The table illustrates the case in which we have discontinuity in a trait G and continuity in another trait H, but discontinuity in the whole genome

As we can see in the table, the trait G alone is sufficient to cluster a population unambiguously, but it is not projectible for the discontinuity of the trait H.

The main point I want to make is that discontinuity does not entail projectibility from one trait to another. In other words, the discontinuity in one trait does not imply a discontinuity in another trait. This situation would be possible only in the case in which the traits inside the genome were correlated in their variation. The table below shows the case of discontinuity in the two traits between P1 and P2. In this case traits are both discontinuous; therefore they are correlated in their variation and the discontinuity in one trait is projectible to the discontinuity in the other trait.

	(G1,H1)	(G1,H2)	(G2,H1)	(G2,H2)
P1	1	0	0	0
P2	0	0	0	1

Figure 5. The table illustrates the case in which we have discontinuity in both traits G and H

In conclusion, in the case of discontinuity for single traits, from a specific genotype of a single individual it is possible to determine with a probability close to 1 the population to which he/she belongs. In this case, clusters that could be constructed on different traits would rarely coincide. Only in the case in which discontinuities in the polymorphic traits were correlated, clusters constructed on different traits would tend to coincide. In this case there would be projectibility from one trait to another and, "by attributing an individual's genotype to a genetic cluster, one would also obtain information on the individual's genome in general" (Barbujani and Belle, 2006).

Consider an arbitrary trait. If there is no discontinuity, then there may still be a difference in trait frequencies between the populations, but in this case, from a specific genotype of a single individual it is not possible to determine with probability close to 1 the population to which he/she belongs.

In an extreme case of continuity, there is likely to be as many genotypes as the number of individuals in a population. Any individual represents a cluster, in the sense that it is a unique combination of genetic traits, which is found only once within a population.

If we look for a realist and unambiguous concept of race, the ontological problem should involve the following questions: Is human genome variation discontinuous? Can we, by attributing an individual to one cluster, obtain information about the whole genome of any individual belonging to that cluster? Are polymorphic traits correlated?

The evidence says that the distribution of different genetic traits is usually continuous across populations and that genetic traits are not usually correlated. Rare cases of discontinuities for single traits with or without correlation could ideally be possible, but only in extreme situations of populations that have been isolated for long time. In general, migration and consequently gene flow has been the main force shaping the human genome. Recent studies on the humankind's genome variation show that populations are characterized by a great continuity in the variation of most traits, making any classification totally arbitrary and meaningless from any biological point of view. For this reason, even if genetic clusters created from statistically significant boundaries among genotypes can be identified, they are not unambiguous and they are not projectible.

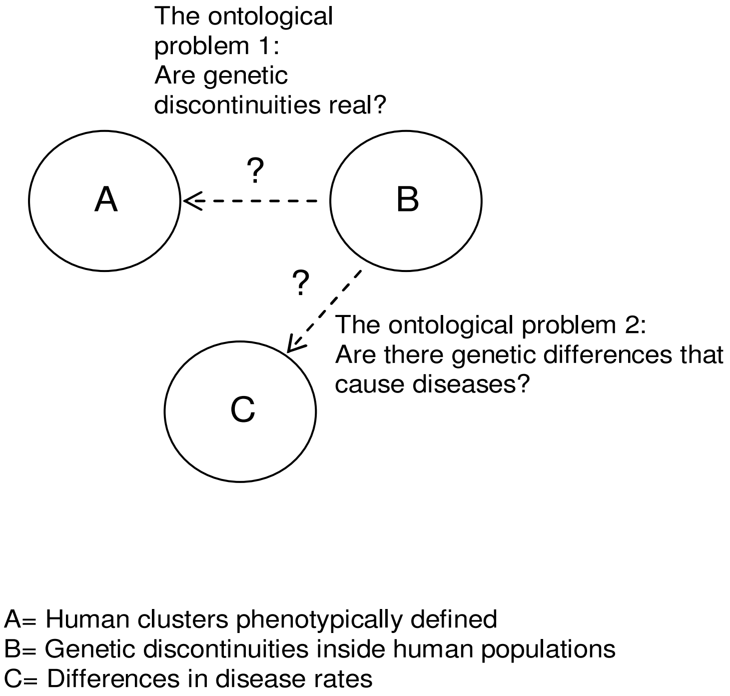


Figure 6. The Ontological Issue

Since here I am not analysing the concept of race in general, but the use of this concept in the specific field of biomedical research, it is worth introducing another ontological issue: the issue regarding the existence of genetic differences among racial clusters that cause diseases.

2.4 The Epistemological Issue: Justification of Races in Biological Explanation and their Projectibility

I am going to analyse the issue concerning whether it is possible to justify the idea that mutations leading to diseases or predispositions to diseases are race-specific.

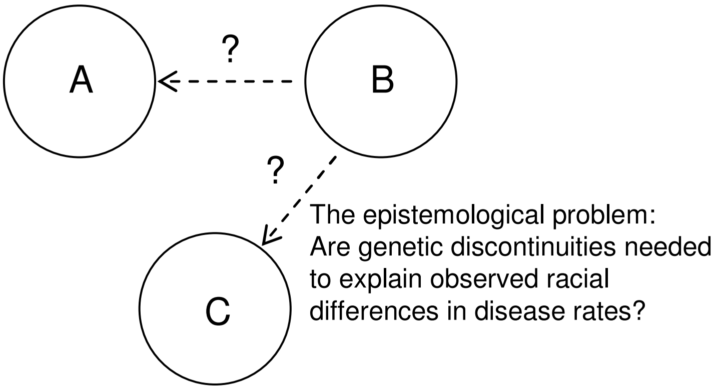


Figure 7. The Epistemological Issue

Consider two clusters: Blacks & Whites. Consider then the case in which there are different genotypes causing different risks of heart disease[5]. The table below indicates that genotypes G1,G3 produce heart disease and genotypes G2,G4 prevent heart disease.

G1,G3	Heart disease
G2,G4	No Heart disease

Consider four different hypotheses. In particular, I shall focus on the first two hypotheses, which I shall name the “strong hypothesis” and the “weak hypothesis”.

---

[5] It is worth noting that by saying that there are genetic causes, I am not denying the existence of environmental causes.

The strong hypothesis says that the genetic factors that cause the disease are discontinuous across racial boundaries. Suppose that blacks have traits G1 or G2, while whites have traits G3 or G4. There is no overlap, so the causes of the disease are discontinuous. In this specific situation, given any individual, and his/her genotype for the disease, one can determine his/her race. The weak hypothesis, however, says that the genetic differences that cause disease are continuous across racial boundaries, that is for example when blacks are G1, G2 and whites G1, G2 with different relative frequencies.

The strong hypothesis implies that races exist, independently of any knowledge of other genetic properties of the populations. The weak one, however, is neutral. Here, I shall not argue that the strong hypothesis is false because races do not exist. I shall argue that the strong hypothesis is unjustified because the weak hypothesis is sufficient to explain the differences in disease rates.

The third hypothesis says that there is no genetic difference between races that is causally responsible for racial differences in disease rates, even though there is a genetic basis for the disease within the whole human population. The correlation between race and disease cannot be explained by differences in genotypes nor by differences in their frequencies, therefore the environment is needed in the explanation.

The fourth hypothesis is similar to the third one, except that it does not assume that there is any genetic basis for the disease, even in explaining an individual's susceptibility to disease. In the two latter hypotheses the correlation between races and disease is explained by a causal relation between environment and disease.

I am going to introduce an idealized example, in which a difference in rate for a specific disease has been discovered inside a human population, between the sub-population of blacks (B) and the sub-population of whites (W).

We observe that 30% of B have heart disease and 20% of W have heart disease. This evidence – that is a correlation between heart disease and racial categories – can be explained with four different hypotheses, represented by the following tables.

B, **G1**	30
B, G2	70
W, **G3**	20
W, G4	80

The strong: genotypic discontinuities between human clusters.

B, **G1**	30
B, G2	70
W, **G1**	20
W, G2	80

The weak: differences in frequencies. Genetic causes are the same (G1,G2), only their frequency changes.

B, **G1**	20
B, G2	80
W, **G1**	20
W, G2	80

No genetic differences, therefore environmental differences must explain differences in disease rates.

B	100
W	100

No genetic causes, therefore environmental differences must explain differences in disease rates.

The numbers in the right columns represent the number of people characterized by the properties of the left column (i.e., B and G1), inside the whole population of 200 people. In the first table on the left, for example, inside the population of 200 people there are 30 blacks with genotype G1, 70 blacks with genotype G2, 20 whites with genotypes G3 and 80 whites with genotype G4.

The strong hypothesis assumes that the genetic traits are discontinuous between the two racial clusters. Here the difference in genotypes between blacks and whites is sufficient to explain the difference in disease rates. The weak hypothesis, instead, assumes only that there are differences in the frequency of genotypes in blacks and whites, but the genotypes are the same inside the two sub-populations. In this case these differences in frequency are also sufficient to explain the difference in disease rates. The third hypothesis assumes that there are neither discontinuities in genotypes nor differences in their frequencies that can explain differences in disease rates. Genotypes can still be causes of that disease for any single individual inside the whole population, but in this case they are not involved in explaining the differences in disease rates between the two sub-populations. For this reason, in the third hypothesis differences in the environment are needed in the explanation. In the fourth hypothesis there are no genetic causes involved in the explanation of the disease, but only environmental ones. This means that in the last case a genetic cause for the heart disease is denied also at an individual level. In the latter two cases, in which differences in disease rates among racial clusters are completely explained by environmental differences, no genetic difference is needed, and a fortiori no genetic discontinuity is needed.

The third hypothesis is usually used to explain cases involving genetic diseases in which a specific genotype is known to be causally related to a specific disease risk. No racial differences are involved in the explanation, since there are only differences among individuals. Note that all the four hypotheses can be used to explain "social" diseases in which individuals belonging to particular social classes and/or doing specific jobs are affected by particular diseases. For none of them exclude the influence of environmental factors. However, the fourth hypothesis can only explain such cases by appealing to environmental causes. While examples explained by the third and fourth hypotheses are well characterized and it is easy to differentiate these two hypotheses from the other two, to differentiate the weak from the strong hypothesis is not so trivial. In order to clarify concretely the difference between these two hypotheses, I am going to discuss a well known disease, the sickle cell anaemia.

2.4.1 The Case of Sickle Cell Anaemia

A concrete example of a disease that can be easily associated with race is the sickle cell anaemia. When I say "associated with race" I mean that differences in disease rates within humans for this disease can be explained by using the strong hypothesis. This disease has a greater percentage in black than in white people, therefore in United States sickle cell anaemia is typically associated with the African-American population. How should this difference be explained correctly? Actually this example is simple, since we have enough information to discriminate between the weak and the strong hypothesis. For we know that there is a genetic basis for this disease and that the specific mutation related to the disease is present in blacks and in whites, albeit with difference relative frequencies. Therefore we know that there is no discontinuity. But it is worth making a general argument anyway. Given that there is a genetic cause with a difference between blacks and whites, we can trivially exclude the fourth hypothesis, since it denies any genetic basis for the disease and the third one as well, since for anaemia there is a difference in the frequencies of the genetic causes between blacks and whites. Finally the decision would be between the weak and the strong hypothesis, but given that the mutation is present also in the whites genotypes, the correct hypothesis is clearly the weak one.

We could still be curious to know why the distribution of this mutation within humans is correlated with the colour of the skin. Why is the mutation for the disease more frequent in blacks than in whites? Looking for the answer we realize that even if there exists a correlation between colour and mutation, the causal relation is actually between environment and mutation. Let me make this clear. The genetic trait for anaemia is found in a higher percentage among people of Mediterranean, Middle Eastern and Indian ancestry. The reason for this geographical specificity of the mutation it is that sickle cell trait offers some resistance to malaria, a very common disease in these regions. In regions where malaria is present sickle cell has been positively selected and consequently the proportion of sick people coming from these places is greater. Since most of black people come from the Mediterranean area, there is a correlation between colour and mutation. In conclusion, sickle cell should not be associated to a particular race, but rather with having ancestors who lived in a malaria-prone region. The proof is that Africans living in areas where there is no malaria, such as the East African highlands, have prevalence of sickle cell as low as parts of Northern Europe.

In this example, it is very easy to understand the fallacious reasoning that bring people to claim a causal relation between colour of the skin and genetic trait. Unfortunately, simple cases like this one are very rare in medicine, where for most of diseases a genetic cause is just assumed and just assumed is also a causal relation between genetic discontinuities and disease rates.

The idea of developing different drugs for a specific disease for different human clusters, in spite of the instrumentalist position of scientists, actually depends on three ontological assumptions: 1) that races exist, that is that genetic variation is discontinuous among racial clusters; 2) that there exists a genetic basis for a disease; 3) that the genetic discontinuity among racial clusters is causally related to that disease. For what would be the sense of developing drugs specific for the racial group of African-Americans, if we did not even know whether the difference in disease rates among racial clusters have actually a genetic cause? Yet, assume that a genotype causally related to a disease is continuously distributed among racial clusters. The difference in frequency of that genotype could be due for example by the fact that statistically most of African-Americans come from a certain environment in which that genotype had been selected in the past. If this is the case, what about African-Americans who have never been in that environment? They would not have that genotype, exactly like other racial clusters. What about African-Americans, with an origin in that environment and living for many generations in another environment? They would probably have lost that genotype.

The third assumption regards the problem of the projectibility of race. By assuming a discontinuous distribution of genotypes among racial clusters and a genetic cause for a disease, there is no justification in believing that those genotypes are causally related to that specific disease. This is the case in which there is a genetic basis for a disease and there are discontinuities in some traits that suffice to cluster unambiguously humans. Does it follow from this information that the genetic cause for that disease is discontinuous among racial clusters? Certainly does not, because as I have said above, discontinuity on some traits does not entail discontinuity for other traits, even if it involves the whole genome. Only a discontinuity with correlation among traits would entails projectibility from one trait to another.

In conclusion, no discontinuity among racial clusters is needed to be causally related to a disease, in the sense that if a certain genotype is distributed in a discontinuous way between two populations, it does not follow that this genotype is causally related to that disease. Therefore, how can the use of clusters built on some kind of genetic differences be justified in the explanation of differences in risk of disease, given that this disease could be caused by other genetic traits? A racial cluster, defined on the basis of several genetic and non-genetic properties, does not have independently testable consequence in medicine. A racial cluster is not carrying any genetic information that can be used in biomedical research in order to make predictions. For all these reasons I claim that the concept of race is introduced ad hoc in the explanation of disease rates.

If there is no genetic discontinuity in the disease causing traits, then there is no point in developing drugs specific to a particular racial group. It would be a marketing ploy rather than scientifically based.

Let us resume the two points that should bring to the “reasonable doubt” the use of any concept of race in biomedical research:

- The concept of race cannot be justified in biological explanation.
- Race does not represent a projectible concept for risk of diseases.

2.5 The *Non Causa Pro Causa* Fallacy

In summary, clusters are defined on the basis of phenotypic, cultural, geographical, genetic properties. Geneticists observe differences in a disease rates among clusters, and clusters are often assumed to be based on genetic discontinuities related to that disease, leading to the following causal fallacy:

Blacks have a greater risk of heart disease than whites → There are genetic discontinuities between blacks and whites that are causally related to this disease

This represents a special case of the fallacy of *non causa pro causa* (literally, “non-cause for a cause”). This fallacy represents a general, catch-all category for mistaking a false cause of an event for the real cause. In our specific case, as shown in Figure 8, the fallacy consists in assuming a causal relation between genetic discontinuities and racial differences in disease rates. In other words, it is a fallacy to postulate the existence of a genetic discontinuity between racial clusters to explain racial differences in disease rates (when there is no independent evidence for such a discontinuity).

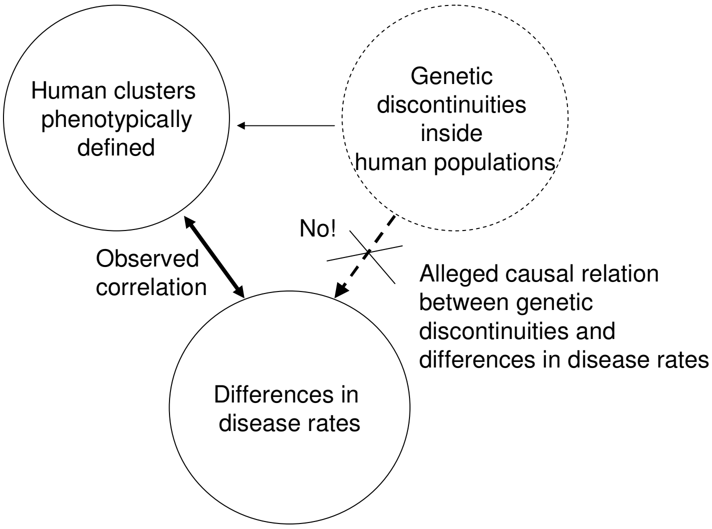


Figure 8. The *Non Causa Pro Causa* Fallacy

I have shown that races are not easily justifiable as causes in biological explanation, while genetic discontinuities are sufficient to explain differences in diseases, they are not needed in the explanation. While the ontological solution of the problem consists in claiming that races do not exist, the epistemological solution I have just offered is to claim that they are not needed and therefore their postulation is unjustified.

We have seen that racial clusters are used as proxies for undetected genetic patterns. This means that, besides the problem of existence of races, there is no evidence for the existence of genetic patterns causally related to those kinds of complex diseases we have considered in our discussion. If there is no evidence for genetic causes, a fortiori there is no evidence for genetic differences among racial clusters. In the case there were this evidence, the weak hypothesis would be the best hypothesis to explain differences in rates among human clusters, since there would be no justification to mention races. This justification is made even weaker in the light of the evidence of very great continuity across human populations.

Only the variability of genotypes among individuals is projectible, and not racial clusters or even races created ad hoc to explain differences in disease rates. For the aim of biomedical research should consist in searching inside genetic patterns related to diseases, and not in assuming racial discontinuities without any projectibility.

The problem of using racial classifications in biomedical research is important because of its ethical consequences in society. Nevertheless, I think that the epistemological and ethical issue regarding the concept of race should be analysed separately, since the usefulness or uselessness of race cannot be established by any argument about the ethical consequences of the use of that concept, but only by means of showing that this concept is not really useful, since it is unjustifiable and non-projectible.

I hope I have also explained why the ontological issue about race has become a dead end; the epistemological approach can finally reach the goal of eliminating ambiguous definitions, obscure uses of racial clusters and the unjustified reification of any kind of racial categorization.

3. Systematics: Classification and Phylogenetic Inference

We have seen that systematics can be used with a different meaning of biological classification. In this case biological classification represents the activity of distinguishing and describing living and fossil species, and organizing those species in a hierarchy, into higher taxa (genera, families, orders, and kingdoms). Phylogeny is the face of systematics that aims to show the historical relationships between a biological groups. To make a phylogeny means to reconstruct the biological history of populations, species and higher taxonomic groups.

3.1 Hierarchy in Classification

The model of biological classification is hierarchical.

All biological classifications are made by subordination and inclusion of classes. All biological classifications are hierarchical and the justification of this structure is that it is considered the only "natural" structure reflecting the real order in nature.

For Simpson (1961, p.13), referring to the zoological taxonomy, there are two forms of classification: the *hierarchy* and the *key* (see Figure 9). A 'hierarchy' is *a systematic framework for zoological classification with a sequence of classes or set at different levels in which each class except the lowest includes one or more subordinate classes*. A 'key' is *a systematic framework for zoological classification (generally used for identification to the exclusion of other purposes) with a sequence of classes at each level of which more restricted classes are formed by the overlap of two or more classes at the next higher level.*

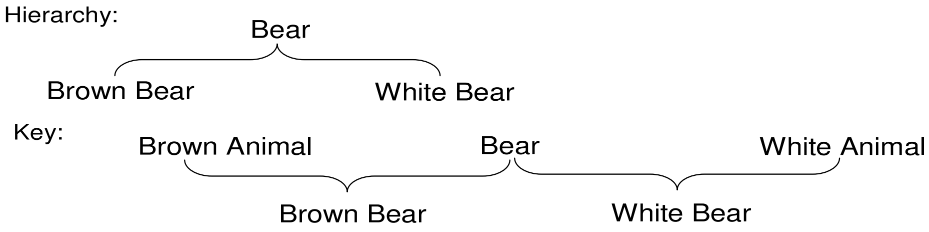


Figure 9. Hierarchy and Key

Hierarchies and keys are distinguished by the fact that in a hierarchy each class belongs to a single class at the next higher level, while in a key each class belongs to two or more classes at the next higher level. Moreover, in a hierarchy the sequence of levels and the subordination of some classes to others is an essential element in the system; on the other hand, a key involves no principle of priority (in the sense that there are not properties more important than others) and has a purely arbitrary sequence. If we look at the diagram in Figure 10 we can notice that “bears”, “nonbears”, “white animals” and “brown animals” are all classes at the same level and “brown bears” and “polar bears” are classes at the next lower level related to those higher classes in a key and *not* in a hierarchic way.

1.

A. Non Bears	Not further classified
AA. Bears	
B. Brown	Brown Bears
BB. White Bears	Polar Bears

2.

A. Brown Animals	
B. Bears	Brown bears
BB. Non Bears	Not further classified
AA. White Animals	
B. Bears	Polar bears
BB. Non Bears	Not further classified

Figure 10. A Key

Since a hierarchy involves principles of priority and these principles are considered derived from real relationships among organisms, a hierarchic classification among organisms is considered a *natural* classification. A key is instead considered *artificial.* Biologists classify species into a hierarchy of groups because a natural classification *must be* hierarchical. There are no *a priori* reasons why a hierarchy should be used instead of a key, but – as Simpson argues – no zoologist would choose the latter in preference of the former. If understanding the psychological reasons (if they are) behind this choice is not very interesting from an epistemological point of view, however it seems interesting to consider the possibility of another way of categorizing the biological reality. We have just seen that there are at least two principles of organizing organisms and we shall see how different criteria (genealogical, morphological) can lead to different ancestor-descendent hierarchies. Note that the *total evidence* (see §6.6

about the principle of total evidence) consists of all organisms and their characteristics. But it is not used in this form. Instead, we categorize organisms by using different principles (hierarchy instead of keys) and criteria (genealogical, morphological) of choosing some similarities of characters instead of others.

Now we can give an operational definition of a hierarchy in biological classification. In a hierarchy, each species is a member of a genus, a family, an order, and so on. The traditional view of classification of species organizes them into a hierarchy of increasingly inclusive groups called also taxonomic ranks: genus, family, order, class, phylum, kingdom.

In standard methods of phylogenetic inference, the reconstruction of trees begins with a classification into groups or clusters. Historical concepts of species and race assume that two individuals from the same group share the same phylogenetic history. However, in the case of race, it is a scientific question whether two individuals from the same racial cluster share the same phylogenetic history, as the standard method assumes. By reconstructing phylogenies from molecular evidence this hypothesis can be tested, and it fails. It fails because migration has been the main evolutionary force shaping the human gene pool. Therefore the historical concept of race is empirically undermined. In the case of species, different species concepts can lead to different phylogenetic trees, therefore in phylogenetic inference phylogenetic species concepts are preferred to biological species concepts[6]. It must be said that this is true whether we use phenotypic data; molecular data, instead, make useless the concept of boundary among organisms and consequently the concept of classification, because in molecular evolution the only meaningful boundary is the boundary among sequences (see §5).

---

[6] See about this topic, Baum and Shaw 1995; Maddison 1995.

3.2 Systematics and Trees

A hierarchy can be phenetic or phylogenetic, in according with which criterion (respectively, morphological or evolutionary) is used order to construct it (see below). For many biologists a natural classification must also be a historical classification. From this perspective, a phylogeny can be the only natural classification. A historical classification is often used as a synonym of evolutionary classification, because these biologists think that after Darwin any historical classification must have an evolutionary meaning. Often, by ‘past’ biologists mean evolutionary past and by ‘phylogeny’ they mean an evolutionary phylogeny. Note that both the morphological and the evolutionary criterion use similarities, but the evolutionary criterion consider only a certain kind of similarity (see below).

A phylogeny is reconstructed under the assumption that the evolutionary process is a branching process. The most used representation of a phylogeny is a *tree*. A tree is made of nodes, branches and leaves. The leaves, aka nodes on the tips of a tree, represent the current taxonomic units, while the internal nodes represent the ancestral ones. The nodes on the tips of the tree are named OTUs (*Operational Taxonomic Units*), and they represent the characteristics of species we can observe and analyze. A tree can be *rooted* and *unrooted*. There are some methods that generate rooted trees automatically and others that do not. When we use the latter ones, we have to fix an *outgroup*, that is one or more OTUs that we know to be older than all other OTUs we are considering in the cladogram or tree.

Nowadays molecular data are often used to reconstruct both species trees and genes trees. Molecular evolution is the branch of biology which is focused on reconstructing trees from molecular data sets. In molecular evolution OTUs must represent the *homologous* sequences, the DNA sequences that two taxa share because they have inherited from a common ancestor.

I shall present here some general ideas about molecular evolution.

“Errors” arising in the genetic transmission over the years shaped the genetic diversity among organisms, from the hypothetical primitive form named LUCA (Last Universal Common Ancestor) to the current living forms. Molecular evolution studies the evolution of organisms by studying the speed and the kinds of errors (I shall use with the same meaning changes or mutations) which arise inside the genetic material and its products, like mRNA, rRNA, tRNA, protein sequences. Evolutionary processes driving the variation in genetic diversity can be considered *indeterministic*. By ‘indeterministic’ process I mean a process where random or directional mechanisms can appear in modifying a framework in which mutations randomly arise. Both these mechanisms are able to change the direction of evolution, in the sense of changing the allele frequencies trough populations and species. At the present time, we do not know when and where random or directional mechanisms arose, arise and will arise. What we are able to know is the final result of these processes in the time, that is new mutations fixed in a population. Errors that initially arose in the gene pool of an individual can spread through a population when they are fixed in it by a random evolutionary process as the genetic drift or a directional one as the natural selection. Errors can be simply substitutions of a nucleotide with another along the DNA sequence. But also other kind of changes can arise, as the delection or insertion of fragments of DNA and various chromosome rearrangments. This explain why organisms, also if they share the same ancestor, they own genomes with very different dimension.

When we use molecular data sets in order to reconstruct species trees, several problems can arise in the complex process of transferring the information from the phylogeny of DNA to the phylogeny of species, and at the end of this process genes trees can fail in reflecting the species trees. For a genes tree, a tree constructed by using homologous genes, is not always identical to the species tree to which the sequences belong. Only if the homologous genes are *orthologous*, that is they are the result of a process of speciation and not duplication, a gene tree will reflect a species tree. Genes that arise for a process of duplication are termed *paralogous* (see Figure 11).

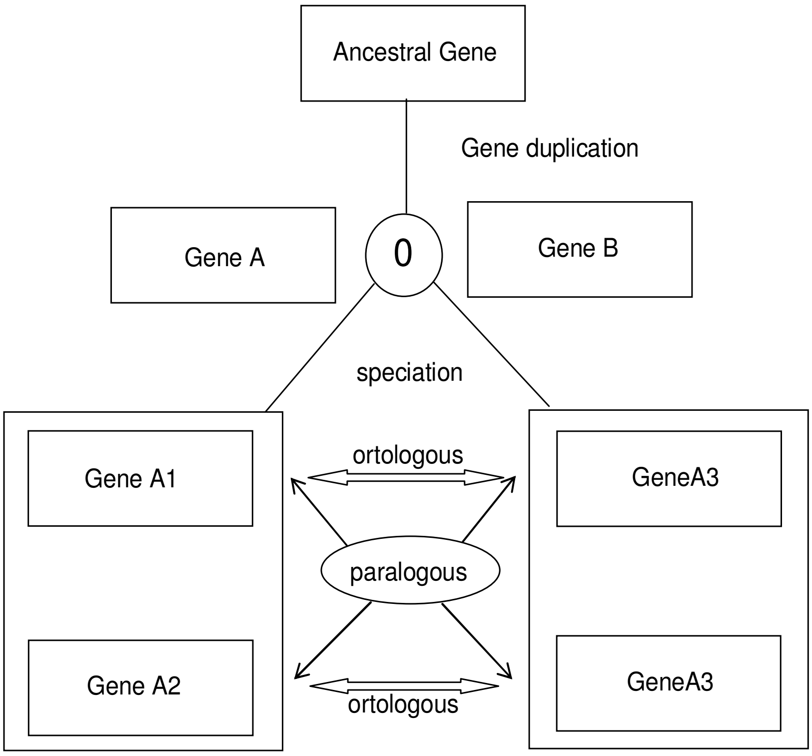


Figure 11. Ortologous and Paralogous

Another kind of similarity is named *xenology*. Two similar gene sequences from two organisms are xenologs when they result from a process of *later (horizontal) gene transfer*[7] between the two organisms. Also for this kind of similarity gene trees will fail to represent species trees.

A tree is the most used representation of a phylogeny, and it is based on the hypothesis that exists a *branching process* where there is an ancestral species A that

---

[7] Animal vectors can carry retroviruses and transfer them among organisms - that can belong to two different species - with the result of a transfer of nuclear DNA.

giving rise to two daughter species, B and C. These two objects could have, in turn, some offspring or could have none. The principle of a branching process is that for each species inside the tree there is only one possible path back to the root of the tree. In other words, for each species there is a unique immediate ancestor. However, from an ancestor we can have more than two descendants. In this case the node is *unsettled* (see Figure 12).

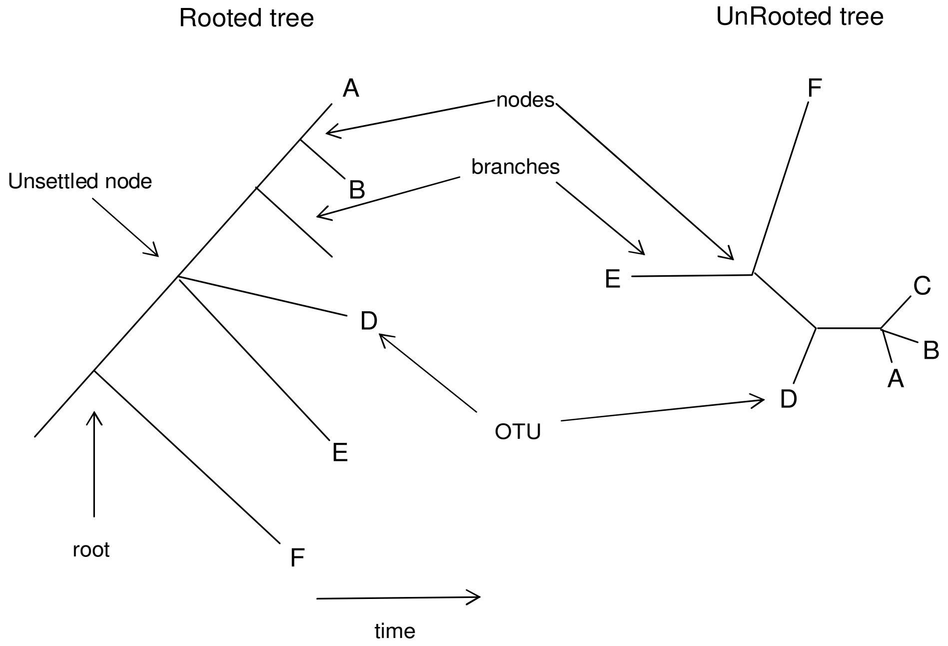


Figure 12. Rooted and Unrooted Trees

3.2.1 The Requirement of Monophyly in Phylogenetic Systematics

In an evolutionary process it is possible to recognize groups of objects called monophyletic groups. *A monophyletic group is a group composed of an ancestor and all of its descendants* (see Figure 13). The concept of monophyly is considered a necessary requirement in phylogenetic systematics and it is used for genealogical definitions of species. Three different concepts use phylogenetic relations in order to classify species: evolutionary, cladistic, and phylogenetic. However, we should note that species *belong to* monophyletic groups, but it is false that each of them *is identical with* a monophyletic group (Sober 2000), because of the fact that species ancestors of other species belong to monophyletic groups but they cannot be monophyletic groups. The requirement of monophyly is appropriate for superspecific (higher than species) taxa but it fails with respect to the species concept itself. This is a problem for the three *phylogenetic species concepts*, in which species are identified with particular segments of a phylogenetic tree, such as lineages of ancestral/descendant populations.

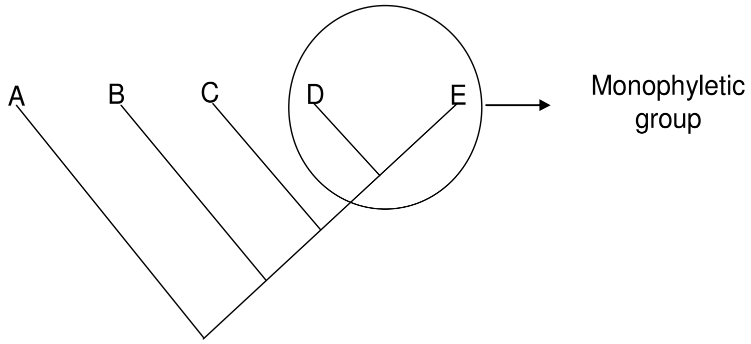


Figure 13

Hypotheses about monophyly are standardly represented in branching diagrams called *cladograms*. Note that a cladogram is not a phylogenetic tree: a cladogram displays taxa at its tips only, whereas a phylogenetic tree has species at its tips, in its interior, and at the root. While a tree implies a cladogram, the converse is not true. Cladograms can individuate monophyletic groups and say something about the relationships between species. Trees individuate such groups and say something about the ancestor/descendant relationships between species. Moreover, a cladogram describes only the phylogenetic relations between different nodes, without giving any temporal significance to the length of branches. In contrast, in a phylogenetic tree the length of branches is considered proportional to the *evolutionary distance* between them. Sometimes in the biological literature, all the phylogenetic representations considered are trees. However, the most informative trees I just described are called *phylograms* and the others cladograms (see Figure 14). In the methods of reconstructing trees that use genetic sequences, the evolutionary distance inside a phylogram corresponds to the genetic distance and it is expressed by substitution/site, where substitution is a particular kind of mutation that is usually considered for studying phylogeny of gene sequences.

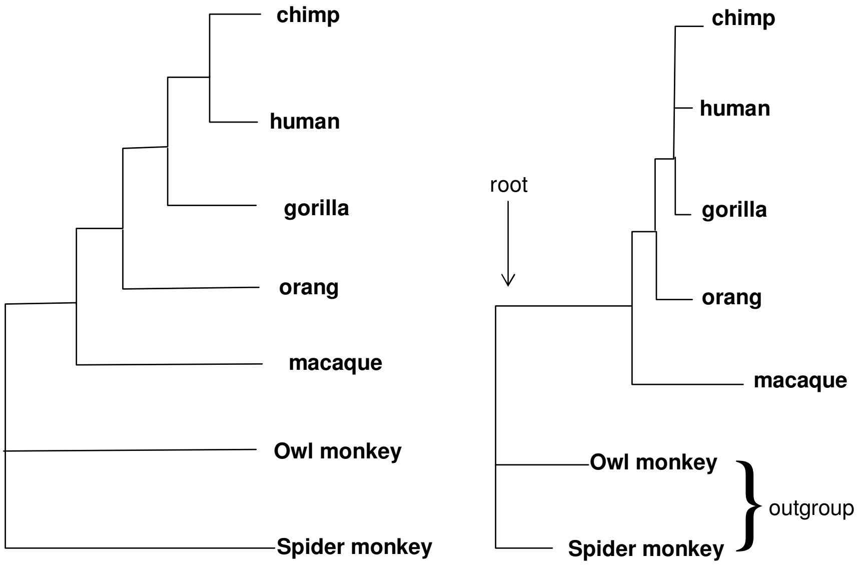


Figure 14. Cladogram & Phylogram

Monophyly is a requirement necessary but not sufficient in order to generate a hierarchical classification (see Sober, 2000). Some other organizing principle is needed to generate a hierarchical classification. For example, a branching structure: a branching structure where taxa are required to be monophyletic describes a sufficient condition for a classification to be hierarchical. In other words, *the set of all monophyletic groups defined on a branching structure constitutes a hierarchical classification*. This is the main idea of the cladistic school of systematics.

3.2.2. Systematics with or without the Evolutionary Theory: Overall Similarity, Cladistics and Evolutionary Taxonomy

We have seen that in order to construct a hierarchical structure, we can use two main different principles of classification: phenetic (morphological) and phylogenetic principles. Let me clarify better these two different criteria. The first one is the principle of pheneticism and it consists in giving a definition of taxa by the *overall similarity* of their members and it has not the aim of reflecting genealogical relatedness. The second gives instead a definition of taxa by using genealogical relationships with the aim of reflecting genealogical relatedness.

Overall similarity is the set of observable phenetic attributes and the phenetic method consists in grouping two species if they look more like each other than either resembles any other species. Finally, a phenetic classification consists of a hierarchy of levels, where the members of different groups at higher levels have decreasingly similar appearances, that is, observable characteristics. In a phenetic classification species are classified only by their similarity, without any consideration of their phylogenetic past. The point is that the phenetic criterion can be considered just as an operational criterion that can be useful in grouping things without the need for knowledge of the past and without a theory of the biological past.

The phylogenetic principle, however, has a historical perspective. So, for classifying things with this principle, we require that these entities have genealogical relations. Following this principle, species are classified according to how recently they share a common ancestor. If two species share a more recent common ancestor we will group them in a lower level than two species sharing a more distant common ancestor, and so on. Finally, we will have the set of all living beings, which contains all descendants of the most distant common ancestor of life. In most cases these two criteria give the same groupings. For example, if we consider the classification of a butterfly, a beetle, and a rhinoceros (Figure 15a), the butterfly and beetle are more closely related both phenetically and phylogenetically. For, the butterfly and the beetle look phenetically more alike and at the same time they share a more recent common ancestor with one other than either does with the rhinoceros. But in other cases phylogenetic and phenetic principles can disagree, or we can say that the phenetic principle fails to reflect genealogical relatedness, because of the fact that the phenetic principle does not always reflect relations of monophyly. There are two main reasons for this disagreement: the case of an evolutionary convergence and the case of a very rapid evolution of two groups of descendants from a common ancestor. In the first case we have, for example, adult barnacles that superficially resemble limpets. So, if we want to classify an adult barnacle, a limpet, and a lobster (Figure 15b) phenetically, we have to put them together, even if the barnacle shares with the lobster a more common ancestor. The second case is well represented by the classification of reptilian groups (Figure 15c). In this case the phenetic and the phylogenetic classifications differ because some descendants of the common ancestor like the birds diverged from their ancestor with a very rapid evolution and left behind quite distant related groups but that resemble one another phenetically like the crocodiles and lizards.

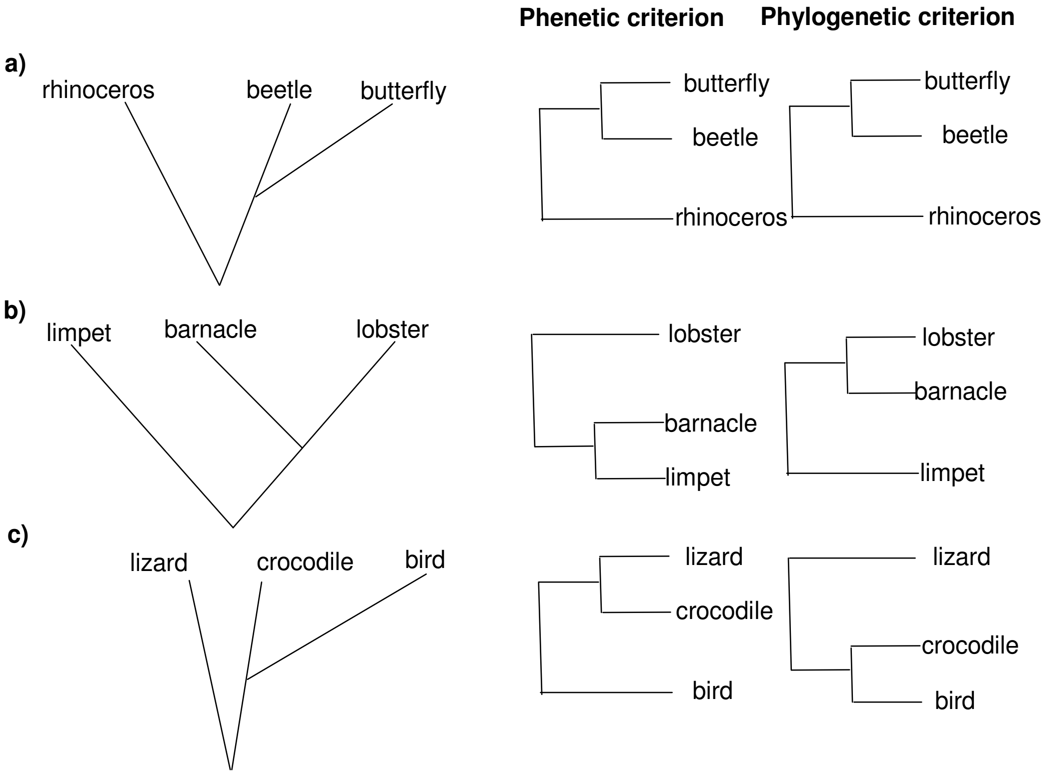


Figure 15

Let us consider now another example of these two cases of disagreement. Consider the three taxa A, B, and C (see Figure 16). At the root of the tree there is the common ancestor and the character state of the ancestor is denoted by 0. In pattern (a) taxa A and B share the same character state 0 that they have inherited unmodified from the ancestor, while the taxon C has evolved a novel characteristic, denoted by 1. The monophyletic grouping is A(BC) and it is not reflected by the overall similarity of the three taxa. In pattern (b) taxa A and B share the same new state denoted by 1. 1 represents a new characteristic that A and B have independently evolved, while C has retained the ancestral state 0. Here again the criterion of overall similarity fails to show the phylogenetic pattern.

(1973).

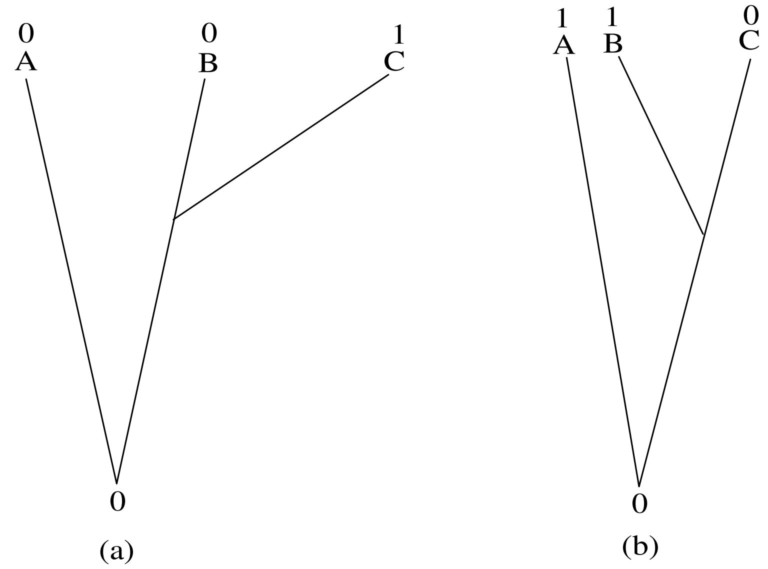


Figure 16

When two taxa share a characteristic because they have inherited it from a common ancestor, this similarity is called *homology*. Otherwise, when the similarity is due to the independently inheritance of a novel condition it is called *homoplasy*. Pheneticists do not care about this distinction, but for them only one type of similarity exists, the overall similarity, without considering the fact that this similarity can reflect homologies or homoplasies.

So, two are the fundamental principles of biological classification, but three are the schools in biological classification: phenetic, cladistic and evolutionary.

The most important school of phenetic classification, based on the principle of overall similarity, is the *numerical taxonomy*, and for this reason in the modern biology the terms phenetics, numerical phenetics and numerical taxonomy are used almost interchangeably. Two names are associated with this school: P. Sneath and R. Sokal (1973).

Cladistic classification is the school founded by the German entomologist W. Hennig

(1966) and his followers. Hennig called it *phylogenetic systematics*, but this school is also known as *cladism*. The school of cladism wants a classification to reflect precisely the evolutionary branching process.

Finally, the third school represents a synthesis of phenetic and phylogenetic methods and is called *evolutionary taxonomy*. This school has been particularly defended by Mayr, Simpson, and Dobzhansky.

If we consider the classification of the *Reptilia*, for example, the phenetic school put lizards and crocodiles together apart from birds, while the cladistic school put the crocodiles and birds together apart from lizards. The evolutionary taxonomy represents a kind a compromise between the two schools. The idea of this school is that similarity can override genealogy but not always. In the case of *Reptilia*, for example, the evolutionary taxonomy put lizards and crocodiles together apart from birds, because it considers homologous the characteristics shared by these two taxa. But, if we consider the case of the classification of Marsupial Wolf, Placental Wolf and Mole (see Figure 17), they do not place marsupial wolves with placental wolves, because the characteristics that they share are not homologies.

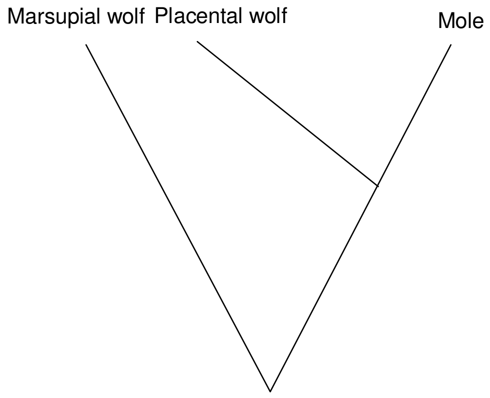


Figure 17

So, when similarity is considered as reflecting homologies, the evolutionary taxonomy classifies by similarity; when the similarity is considered not to reflect homologies, it classifies by genealogy.

3.3 The "Best" Classification

Now we have seen the three different school of classification and we can ask the question: What is the "best" classification? In the sense of asking what classification can better explain the diversity in traits among species. The more plausible, the more reliable. If we use theories of probability (see par.2) in order to make a decision about the classification that should be used, the "best" is the more probable – in a Bayesian framework – or – in a likelihoodist framework – the one that is better supported by the evidence. If we think that behind a probability there is something more than a degree of belief or confidence, or in other words we reject the subjective view of probability, we can think that the probability of a hypothesis is correlated in different ways with its plausibility and reliability. We shall see how in phylogeny we can apply this idea.

We have seen that each different systematic school has different goals. Phenetics focuses on overall similarity, cladistics on branching pattern, and evolutionary taxonomy on a mixture of branching pattern and adaptive similarity.

If we regard these different hypotheses as hypotheses that compete with each other, we should ask: how can we decide among these different philosophies of classification, if they have different goals? We could want to regard the choice of systematics as a matter of arbitrary convention, considering all the classifications as arbitrary, but we could also look at them as non-arbitrary conventions in Poincaré's conventionalism (see above).

We have seen before that the phylogenetic classification is considered by many biologists to be the only natural and non-arbitrary method. On the other hand, pheneticists claim that a classification based on the overall similarity is less arbitrary than a classification based on cladistic or evolutionary criteria, because it is "theory neutral", in the sense that it is based on a direct and pure observation of natural characters and is not influenced by theory, in particular by evolutionary theory. For, while evolution is considered a necessary assumption of phylogenetic classification as is cladistic and evolutionary taxonomy, it is not considered necessary for the phenetic classification, because here the organisms are classified by similarity of appearance, independently of their evolutionary relations. They claim that by using the phenetic criterion we can construct an "objective" classification that requires no theory at all, but that consists in the simple act of observing traits and recognizing defined and named groups. A group, in this simple sense, is a collection of organisms that share a particular defining trait. Sober (2000) pointed out that also the overall similarity criterion needs assumptions about evolution. For, a classification made by using this criterion is *consistent* – in the sense that it converges to the true classification for an infinite number of characters – only under the assumption of uniformity of rates along the branches. Uniformity of rates means that we impose the same rates of changing of character to branches within the same temporal period. In terms of probability, we can express this idea by saying that the probabilities of changes for each branch of a tree are the same within the same temporal period.

The *condition of adequacy* says that the method of overall similarity will not fail and will converge on the true phylogeny if we consider a large data set. And the uniformity of rates is needed for that condition of adequacy to be satisfied.

We can finally ask: if the method of overall similarity satisfies the condition of adequacy only in a world where the rates are uniform and we know that in this world this is false, can we still use this method? The answer depends on what we are looking for: truth, utility, accurate predictions? Also there is no method in phylogenetic inference that does not need some assumptions about the evolutionary process, and most of them are known to be unrealistic (see chapter 3).

So, assumptions about the evolutionary past are necessary in order to classify species phenetically, if we want this method to satisfy the principle of adequacy and to be (at least with an ideal infinite data set) something more than an instrumentalist criterion. If we are not interested in the true classification, we can keep using the phenetic criterion without any consideration of the evolutionary past.

Let us analyze now the question of the objectivity in numerical taxonomy, considering the group of Chordata, which is defined as the group containing all animals that possess a notochord. Also if in appearance it can seem simple and objective to create this classification grouping animals that share notochord, it is easy to understand that this classification is not objective. For, if it is true that a biological classification is only grouping living beings on the basis of observed characters, it is also true that for different traits chosen we can have different classifications. So, the question is: why finally decide to consider some characters and not others, obtaining one classification and not the alternative ones? On what basis can such a decision be made? The trait selection inevitably needs some criterion. So, it is evident that observational data are influenced by theoretical considerations about what kind of character we want to consider in our classification. Ridley stresses this point showing that Chordata is not the only possible grouping, but it is the output of a subjective choice:

> Chordata happens to be a group that is generally recognized, but by the same method we could define other groups that are not normally recognized. We might, for instance, define the Ocellata as the group of all living things that possess eyes. It would contain most vertebrates, many insects and crustaceans, some molluscs, and worms, and some other odd invertebrates. The Ocellata has not, so far as I

> know, ever been considered as a taxonomic group; but, if classification is only a matter of defining and naming group, we might ask why it is not (Ridley, 1985, p. 73).

So, if eyes define Ocellata, notochords define Chordata. How can pheneticists defend their school from an accusation of subjectivity with respect to choice of traits? The answer is: with cluster statistics. Given many measurements of many traits in the units we want to classify, cluster statistics permits the formation of groups (or clusters) of units according to their similarity in all traits, aggregating the units with the shortest distances (the distance in similarity, there is the difference between the values of trait in two groups) to each other. The way of forming these clusters is hierarchical, in the sense that a hierarchy clusters as more and more distant units are added in. Clusterists believed that because of the huge number of traits used in the cluster statistics, the groups discerned by the cluster statistics will be less arbitrary. In other words, it is believed that statistics could make the phenetic method less arbitrary because of the fact that using the same statistical techniques, everyone could be able to obtain from the same animals and plants, the same classification. The objectivity and repeatability that numerical taxonomy claimed for itself is illusory, and the reason is relatively simple: there is in fact more than one way of clustering groups. By choosing different methods of clustering we obtain different hierarchies of groups. Since the principle of overall similarity claims that taxonomy should found its objectivity on the use of a technique without any theoretical support, basically there is no possible theoretical justification behind the choice of different statistical criteria. As Ridley points out,

> The principle of numerical taxonomy provides no guidance among the different cluster statistics. It implies no criterion by which to choose among the different hierarchies produced by different statistics. The principle of numerical taxonomy is

> to classify according to "overall morphological similarity", but overall morphological similarity can only be measured by a cluster statistic. There is no higher measure of overall morphological similarity against which the different cluster statistics can be compared. When different statistics conflict, the practical numerical taxonomist has to decide which he prefers. He can make a choice, of course; but it will have to be subjective (Ridley, 1985, p. 81).

In accordance with what has been said about cladism and phylogenetic classification, it should be evident that biologists holding the cladistic principle use the term "phylogenetic classification" as a synonym for a phylogenetic tree. Consequently, if we know the phylogenetic tree for a set of taxa, this information is considered sufficient to make a classification. Instead, this information is worthless for a classification based on phenotypic criteria.

4. Cladistic Parsimony and the Philosophical Problem of Parsimony

In the school of cladism, the fundamental assumption is that there is only one correct phylogeny, that is the hierarchical reconstruction of the real relationships among living beings.

> The phylogenetic tree is a unique hierarchy. It really is true, of any two species, that they either do or do not share a more recent common ancestor with each other than with another species. In phylogenetic classification, there is no problem of subjective choice among different possible hierarchies. *There is only one correct phylogeny* (*Ivi*, pp.81-82. Italics added).

In phylogenetic systematics the cladistic method is also called *cladistic parsimony*.

The goal of cladistic parsimony is to know, for each species, with which other species it shares its most recent common ancestor. In order to discover it, this method proposes to look only for traits that are "novelties". According to whether a particular trait is an earlier or a later evolutionary stage in evolution, it can be called a *primitive* (*original*, *ancestral*, *plesiomorphic*) or a *derived* trait (*apomorphic*). Obviously, a trait is not an absolute primitive or derived, but it depends on which other stage it is being compared to: it can be primitive with respect to later stages but derived with respect to earlier ones. The ancestral similarities are named *symplesiomorphies*, and the derived similarities, *synapomorphies*.

In order to reconstruct a phylogeny, the cladistic method considers *only* similarities between the apomorphic traits, the synapomorphies. The phenetic method, instead, consider all the similarities and give to any similarity the same weight.

How do we decide what character is ancestral and what derived? In order to obtain this information about the state of character to know each character's *polarity* is necessary. *Polarity* of characters is defined as the state of characters used in a cladistic analysis, either original or derived. Original characters are those acquired by the older ancestor in the phylogeny (the outgroup or root of the tree). Derived characters are those acquired by the most recent common ancestor of the taxa under consideration. To infer the genealogy of the taxa in a particular group by using the method of cladistic parsimony, to polarize the characters is necessary. For this purpose in unrooted trees we must chose an *outgroup* (see above) and use the character state of this outgroup for estimating the ancestral character state. This method used to determine the polarity of the characters is called *outgroup comparison* and it is shown in Figure 18.

> Cladistic parsimony, […], is an elaboration of the idea of *special* similarity. It is not just any kind of matching that counts as evidence of common ancestry; it is derived-not ancestral-similarity that has evidential meaning (Sober, 1988, p.27).

Consider for example three taxa A, B, C characterized by one character in two states: *t* and *t'*. In order to decide what is the ancestral state and what the derived, we want to look at the state of the character in the species outgroup, in this case D. If the species D has *t'*, *t* is a plesiomorphy, if the species D has *t*, *t* is an apomorphy.

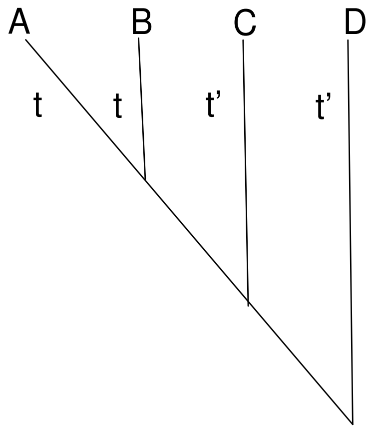


Figure 18

By considering only some particular similarities of the total similarities between taxa, cladistic method brings often to choose different profiles from the phenetic method. For example, let consider three taxa A, B, C and 10 different characters, where "1" describes the derived status, and "0" the ancestral status of each character.

Characters	Species A	Species B	Species C
1-8	0	0	1
9-10	0	1	1

In this case cladistic would consider only the similarity of the two characters 9 and 10, and so it would decide for the profile A(BC), also if they share only two character on 10! The phenetic method, however, would consider all the similarities, and finally it would decide for the profile (AB)C, since A and B share 8 similarities and B and C only 2.

5. Biology without Classification: a “Molecular” Way of Thinking

So far I have introduced several issues about biological classification. It has been claimed many times that classifications with a historical meaning are needed in phylogenetic inference, since other classifications could bring to distort evolutionary histories. In other words, when species are to be taxa used in phylogenetic inferences, we need a properly phylogenetic species concept. However, I claim here that phylogenetic reconstructions - if based on molecular sequences - do not need a biological classification and *a fortiori* a specific kind of classification. In molecular evolution the evidence to be interpreted in an evolutionary sense is given by the comparison among DNA and aminoacidic sequences. Molecules considered to be essential for life are chosen to be compared, since they carry a memory of the history of the organisms; these molecules have been more conserved during the past centuries with a very few number of variations. To reconstruct a phylogeny without a classification is possible and it involves the recognition and valuation of boundaries between molecules. It is worth saying that this “molecular” way of thinking in evolutionary biology does not aim to be reductionist. For I am not denying the biological value of thinking about individuals and populations as units of evolution differently adapted to their environment, which is the founding idea of species concepts like the ecological and biological ones and race concepts like the concept of “ecotype”[8]. I want just to point out that from a methodological point of view, when we focus on the historical meaning of the relationships among organisms, to work with a rigid classification in species is not

---

[8] About the concept of race as “ecotype” - that is a local population adapted to particular environments - see Pigliucci and Kaplan 2003.

needed anymore, if we look at the molecular sequences. For what concerns the problem of race, I argue that the only possible definition of race with a phylogenetic meaning is that one concerning the groups of mutations and their variation in the time. The history of *Homo sapiens* is not represented by a history of races, instead it is represented by a history of mutations that vary in space and time over the centuries. The concept of race as a group of human populations has a sense only as a geographical and adaptive concept concerning a very limited set of phenotypic differences (i.e. the colour of the skin).

6. Probabilistic Theories

Phylogenetic systematics is a discipline founded on probabilistic theories. Our predictions in the past are probabilistic and this means that it is possible to know the past by means of different hypotheses which have a different probability to be true. The fact that our epistemology of the phylogenetic past is probabilistic does not force us to believe that the process linking the past to the present is probabilistic, it does not force us to have a metaphysical commitment on the nature of the evolutionary process.

In this section I am going to introduce some concepts of two different probabilistic frameworks, which are used to justify our beliefs in hypotheses about biological past. There is a fundamental presupposition behind any application of probabilistic approaches to the reality and it is that scientific reasoning is essentially a reasoning in accordance with the principles of probability. According to the probabilistic view, scientific reasoning should be conducted in probabilistic terms.

Probability can be interpreted as our degree of belief or as a relative frequency. In the first case it can be defined as our degree of confidence in something uncertain to happen. This idea belongs to the *subjective view* of probability. B. De Finetti is one of the main names tied to this particular view. For him, probability is always subjective, it is a question of feeling and its use in making decisions or predictions always implies taking responsibilities.

6.1 Bayesian Philosophy of Science

Bayesianism is not the same as Bayes's theorem. The theorem states that the conditional probability P(H|O) - the probability of H, given O – is a function of the three quantities P(H), P(O), P(O|H). The theorem says that:

$$P(H|O) = P(O|H)P(H)/P(O)$$

For a Bayesian, evaluating an uncertain claim should be driven by calculating the probability of the claim in the light of given *information* (O) that is, calculating the *posterior* probability of the claim. Posterior probability is considered by Bayesians as a way of measuring the *plausibility* of a hypothesis. P(H) is the prior probability of H, the probability that H has *before*[9] the event E. P(E|H) is the *likelihood* of H. The likelihood of a hypothesis is the probability of the total observed data, given the hypothesis (not to be confused with the probability of the hypothesis given the data). In other words, it is the probability that the hypothesis H confers on the observation O. Likelihood can also be thought of as a degree of fit of the hypothesis with the data.

The theorem is derivable from a definition of conditional probability. This is the Kolmogorov's (1950) definition of it:

$$P(H|O) = P(H\&O)/P(O)$$

---

[9] Note that *prior* in this context means *before* the evidence O, and it does not mean independent from any experience.

Consider H=A and O=B, so that we have:

$P(A|B)=P(A\&B)/P(B)$

Now, switching A and B in the equation, we obtain:

$P(B|A)=P(A\&B)/P(A)$

From the two equation we can see that the probability of the conjunction A&B can be expressed in two equivalent different ways:

$P(A\&B)= P(A|B)\ P(B) = P(B|A)\ P(A)$

It should be easy to see that from this last equality we can obtain Bayes's theorem.

Bayesianism is a philosophical position, not a mathematical theorem; it is an epistemological framework, in the sense that it uses the probabilistic relationships expressed in the theorem to represent the relationships between theories and evidence. In this framework, the connection between a hypothesis and the evidence is measured by two different kinds of probability, in the posterior probability and in the likelihood. But both hypothesis and evidence have also their own unconditional probability, respectively P(H) and P(O). The unconditional probability of the hypothesis is called the prior probability of the hypothesis, that is the probability that you assign to the hypothesis depending to a background knowledge (which can be empirically based), before having observed O. Also the probability you are going to assign to the event O depends on what you have observed before and of any kind of information you have in your background knowledge.

The concept of confirmation of a hypothesis can be expressed in terms of its posterior probability in this way:

1. O confirms H if and only if $\Pr(H|O) > \Pr(H)$
2. O disconfirms H if and only if $\Pr(H|O) < \Pr(H)$
3. O is confirmationally irrelevant to H if and only if $\Pr(H|O) = \Pr(H)$

Thus, if O confirms H, $\Pr(O|H)/\Pr(O)$ will be greater than unity. This view of confirmation expressed by the inequality 1, in which you compare the posterior probability of a hypothesis with its prior probability, can be called the incremental view of confirmation. Let us consider a special case in which H deductively implies O, so that we have that $\Pr(O|H)=1$. In this case, $\Pr(O|H)/\Pr(O)$ cannot be less than unity. Because from the Bayes's theorem we know that $\Pr(O|H)/\Pr(O) = \Pr(H|O)/\Pr(H)$, it follows that $\Pr(H|O)/\Pr(H)$ cannot be less than unity. This means that when you deduce a prediction and the prediction comes true, O is not able to disconfirm the hypothesis. There is a case in which $\Pr(O|H)=1$ and O will fail in confirming H. Exactly when $\Pr(O)=1$. If you know that O will be true before you made that observation, observing O cannot confirm H.

We can also notice that the first proposition is logically equivalent to the following proposition:

O confirms H if and only if $\Pr(O|H) > \Pr(O|\mathit{not}H)$

We can ask what *not*H represents. *Not*H is called the catch all hypothesis. It looks like a simple hypothesis, exactly like the hypothesis H. But actually notH is a composite hypothesis, a disjunction of all the possible alternative hypotheses to H. There is this ambiguity because when we look at the formula we think about logical-mathematical relationships and objects. However, Bayesianism is an epistemology, so when we see *not*H we must think about it not as an abstract symbol. We must think about what it does represent. In *not*H all possible alternative hypotheses are considered, also hypotheses not been formulated yet. Because it is a composite hypothesis, Pr(O|*not*H) cannot be well defined in Likelihoodism (see below).

6.1.1 The Posterior Probability as an Absolute Measure of Bayesian Confirmation

Another way of considering confirmation in Bayesianism is not via comparison between the posterior probability of a hypothesis and its prior probability. This way considers the absolute posterior probability of a hypothesis P(H|O). This view is used when you want to compare two different hypotheses in the light of the same evidence. In this case Bayes theorem shows the comparison of the absolute posterior probabilities of the two hypotheses:

$$\frac{P(h_1 \mid O)}{P(h_2 \mid O)} = \frac{P(h_1)}{P(h_2)} \times \frac{P(O \mid h_1)}{P(O \mid h_2)}$$

It follows that:

P(H1|O) > P(H2|O) if and only if P(O|H1)P(H1) > P(O|H2)P(H2).

6.2 The Problem of *A Priors* in Bayesianism

The most common objection against Bayesian epistemology concerns a prior probabilities of hypotheses. The question is: On the basis of what kind of information should these priors be assigned? Bayesians have two ways to answer this criticism. One subjective and one objective. The subjective is founded on the idea that probabilities describe our degree of belief. If I have a degree of belief in a theory before seeing the evidence of it, this will determine what my prior probability about it is. The problem is that my degree of belief can be different from the degree of belief that another person has about the same theory. If probability describes subjective degrees of belief, it is not possible to show which value of probability is correct and which is not. Bayesians reply to this by saying that once you start to have a certain amount of evidence available, you will obtain nearly the same posterior probability, also if you started with very different priors.

The objective way is founded on the principle of indifference. The principle of indifference (also called principle of insufficient reason) states that if there are $n > 1$ mutually exclusive and collectively exhaustive possible hypotheses, and if they are indistinguishable, in the sense that you have no reason to assign to them different probabilities, then each hypothesis should be assigned a probability equal to $1/n$. Also if this principle is considered the "objective" route of priors assignments, it is meaningful only inside the subjective interpretation of probability, because it is easy to see that the "having no reason" is again a subjective statement!

There are some cases in which the use of the priors is legitimate, that is when the priors are fixed on the basis of an empirical information. This is true also for the priors that are introduced in the calculation of the likelihood of a model (see below). A simple case can be the prior probability of getting a king of spades by picking up randomly a card from a full and normal deck of cards. Its numerical value is 1/52 and it is objective in the sense that it comes from an empirical information about the chance process and it is not obtained from the Principle of Indifference, neither does it reflect your degree of belief.

6.3 Likelihood of Models: Bayesianism vs Likelihoodism

Now we can consider how Bayesianism works with composite hypotheses. A composite hypothesis, also called a model, is usually a disjunction of simple hypotheses. The likelihood of a model is an average and this average cannot be well defined without giving a value to special priors which are the conditional probabilities of every single hypothesis of the model, given the model. When we are working with simple hypotheses, the likelihood is independent of the priors. However, when we are working with composite hypotheses, the likelihood depends on the priors. In other words, while a simple hypothesis confers a well defined probability to an observation, a composite hypothesis does not and you need priors. These priors, disguised as conditional probabilities, are probabilities of values of *adjustable parameters* of the model. For example, consider the model *M*, which is an infinite disjunction of simple hypotheses.

$M: \ x = \theta$ where θ is an adjustable parameter of the model.

Any simple hypothesis in the model is obtained by assign a particular value to the adjustable parameter. For example, the following hypotheses are simple hypotheses in M.

$h_1: \ x = 1.0$

$h_2: \ x = 1.1$

M can be expressed by the following formula:

$M = h_1 \vee h_2 \vee \ldots$

In this example, the priors can be expressed as the conditional probability of the simple hypothesis hi (i=1,2,…), given the model M:

$$P(h_1|M) = P(\theta = 1|M)$$

$$P(h_2|M) = P(\theta = 1.1|M)$$

There is a correspondence 1:1 between the hypotheses in the model and the numerical assignment to the parameters:

Model (M) + values of the parameters → hypothesis (hi) of the Model

where the hypothesis logically entails the model.

Let consider another example, in which we have simple and composite hypotheses about the same given information. Suppose you have a population and you are considering in particular a family in this population. You know that the offspring of this family has A as blood group. This is the given information. Now you want to analyse the likelihood of three different hypotheses: The hypotheses of the blood group of the parents, the hypothesis that the blood group of the mother is A, and finally the hypothesis that the blood group of the father is A. The likelihoods of the hypotheses of the blood group of the parents, relative to the observation of the offspring's blood group, are:

Pr(offspring is A|Parents are A and 0)=1/2
Pr(offspring is A|Parents are A and B)=1/2
Pr(offspring is A|Parents are B and 0)=0
Pr(offspring is A|Parents are 0 and 0)=0
Pr(offspring is A|Parents are A and A)=1.0
Pr(offspring is A|Parents are B and B)=0

Pr(offspring is A|Parents are AB and B)=0

Pr(offspring is A|Parents are AB and AB)=0

Pr(offspring is A|Parents are AB and A)=1/2

Pr(offspring is A|Parents are AB and 0)=0

Hypotheses about the blood group of the parental pair are simple because they have a well defined likelihood. However, the hypothesis H1 that the offspring's mother is A is composite. The hypothesis H1 is a disjunction of simple hypotheses – the mother is A and the father AB, the mother is A and the father A, the mother is A and the father 0, the mother is A and the father B. Its likelihood is an average:

Pr(offspring is A|mother is A) = Pr(offspring is A|mother is A & father is AB)Pr(father is AB|mother is A) + Pr(offspring is A|mother is A & father is B)Pr(father is B|mother is A) + Pr(offspring is A|mother is A & father is A)Pr(father is A|mother is A) + Pr(offspring is A|mother is A & father is 0)Pr(father is 0|mother is A)=

(1/2)w1 + (1/2)w2 + (1)w3 + (1/2)w4 (where w1+w2+w3+w4=1.0)

Another composite hypothesis is that the father is A, and I shall call it H2. Alternative composite hypotheses can have different adjustable parameters, and they can have different likelihoods. The likelihood of the father will look exactly as that one of the mother, except for the fact that the conditional probabilities (adjustable parameters) will be: Pr (mother is AB|father is A), Pr(mother is B|father is A), etc.

Adjustable parameters of the model can be estimated empirically if you know that mating was random inside the parental population. You can estimate the unconditional probabilities by looking at the relative frequencies of the groups ABO in the parental population (rf(A), rf(B), rf(AB), rf(0)). Very often in the scientific practice you do not even have this information about the processes that have produced data you want to consider in your statistical analysis. Your information about whether mating was random in the parental population can be lacking and in this case these weighting terms cannot be specified empirically. In this case the adjustable parameters may be called *nuisance* parameters. If you do not know anything about the mating process you have different choices. You can 1. Impose a prior probability distribution over values of the parameters as reflecting one's subjective degree of belief. 2. Give a value to nuisance parameters by finding the values for the parameters that maximize the likelihood. If you consider the composite hypothesis H1, the values that maximize the likelihood are: w1=0, w2=0, w3=1, w4=0. The hypothesis with the Maximum Likelihood is often named LH1.

It should be noticed that the assignment of these values is based on no additional empirical information. The question is: How can you justify the use of a hypothesis, which seems to be created for the precise purpose to have the Maximum likelihood?

The answer can be: by making assumptions! You can assume a process in which these value of the adjustable parameters are justified. You can assume a "special" process, in which fathers A always mate with mothers A, but mothers A do not chose always fathers A. Under these assumptions, the likelihood of the hypothesis LH1 has a likelihood of unity and so LH1 is more likely than H2. But there is a problem, and it is that now you are not comparing H1 with H2. However, now you are comparing the hypothesis LH1, which is simple (0 parameters) with the composite hypothesis H2. The likelihood of the LH1 is greater than the likelihood of H1 and not equal to it! Summarising, here you consider the likelihood of your bettered simple hypothesis LH, instead of considering the average likelihood of your composite hypothesis H. This strategy for handling nuisance parameters is sometimes called the Best-Case strategy (see Sober, 2002a).

Consider now a case in which you have the information that mating was not random in the following way. The offspring population represents the result of a freak experiment of eugenics, in which males and females in the parental generation with the same blood group have been forced to come together to reproduce. This information would make everything different, because in this case the number of adjustable parameters would be 0. The hypotheses we were comparing before would be simple and their likelihood would be well defined and exactly the same Pr(offspring is A|mother is A) = Pr(offspring is A|father is A) = (1/2)0 + (1/2)0 + (1)1 + (1/2)0 = 1.00. The background information is very important because it can always reduce the number of adjustable parameters of a model in an genuine (objective) way, or at least justify the artificial reduction of parameters obtained by making assumptions more or less ad hoc, more or less unrealistic. For this reason sometimes the likelihood of composite hypotheses can be expressed as Pr(O|H&A), where 'A' represents the background assumptions that influence the determination of parameters.

This was a simple and artificial example, but it shows what really happens in the everyday scientific practice. We shall see that this strategy is adopted by the homonymous method in phylogenetic analysis. In that case you have different hypotheses about different topologies of trees; the evidence about the living species; you do not have any kind of empirical information about the evolutionary process which generated the evidence you observe and any kind of information about the status of characters of the ancestors in any monophyletic group you wish consider. Phylogenetic hypotheses are composite and so they do not have a defined likelihood. They are infinite disjunctions in which each member is a topology with a well defined set of fixed parameters. The parameters in the models cannot be specified empirically and so they are nuisance parameters. If you are a Bayesian you can calculate the average and you need to give a value to the priors P(Hi|M). You can instead use the maximum likelihood method to estimate the nuisance parameters and you obtain the maximum likelihood topology of tree given the information about the status of characters (DNA sequences if you work with DNA) in the current species. We have seen that when you use the maximum likelihood approach, you do not compare the likelihoods of the different models but the likelihoods of the L(M), that is the likeliest special case (phylogeny) in the model. But there is another problem here. If you use the best-case strategy for a model, you can make every possible phylogeny of the model having a likelihood equal to 1, for every distribution of characters. This happens when you are considering a fully realistic model of the evolutionary process. Consequently, inside a realistic process, to discriminate between different phylogenies is impossible (see Sober, 2002a). You need to use some constraints (assumptions) about the process of change. Constraints can involve the frequency of changes (mutations) of characters between branches; the independency of change of different characters; the fact that different characters follow the same rules of change and have a constant rate of change; features of the tree topologies (i.e. length of branches); kind of changes (i.e. parallel, convergent, multiple; transition, transversion). You can have different models with a different number of adjustable parameters (constraints), where simpler models (with a smaller number of adjustable parameters) entail more complex models. Then you are able to use the maximum likelihood approach in order to discriminate between phylogenies, because every phylogeny will have a different maximum likelihood in different models of processes. Are these constraints realistic? Obviously No, and so neither is a constrained model. But if you want to use a realistic complex model you cannot use the maximum likelihood approach for inferring phylogenies.

6.4 The Posterior Probability of the Model

In a Bayesian framework, the calculation of the posterior probability of a model involves two different kinds of priors: one kind of prior is the prior probability of the model (that is the same we have for single hypotheses) and the other one is the probability of the simple hypotheses given the model. Let us consider, for example, a model that is a disjunction of two hypotheses h1 and h2:

$$M = h_1 \vee h_2$$

When I compare the posterior probability of two models, I shall have:

$$\frac{P(M \mid O)}{P(M^{'} \mid O)} = \frac{P(M)}{P(M^{'})} \times \frac{P(O \mid M)}{P(O \mid M^{'})}$$

where $M = h_1 \vee h_2$ and $M^{'} = h_3 \vee h_4$

The likelihood of a model is:

$$P\left(O \middle| M\right) = pP\left(O \middle| h_1\right) + qP\left(O \middle| h_2\right).$$

The priors are:

$$p = P\left(h_1 \middle| M\right) \text{and } q = P\left(h_2 \middle| M\right).$$

6.5 Likelihoodism

Likelihoodism could be thought to as Bayesianism without *priors*, in the sense that the share the idea of likelihood and the principle of total evidence (see below). We have just seen that in a Bayesian framework, evaluating an uncertain claim should be driven by the posterior probability of the claim, which introduces prior probabilities and likelihoods. Instead, in a likelihood framework, you do not consider the posterior probability of a hypothesis, but only the likelihood of simple hypotheses, that is the probability that a hypothesis confers to data. This idea is also expressed as how much the data favor or support a hypothesis. Moreover, if you chose the likelihood framework, you do not need the prior probability of the hypothesis you are evaluating. The epistemology of Likelihoodism is considered by likelihoodists a more "objective" epistemology in comparison with Bayesianism, for two reasons. First, when you decide to use only likelihoods and no priors you avoid the need to judge in probabilistic terms all the situations in which you have to give priors. Second, likelihoods are in themselves considered more objective than prior probabilities, or posterior probabilities. Royall (1997) expressed this idea by affirming that likelihoods do not tell you what to believe or how to act or which hypotheses are probably true; they can only tell you how to compare the degree to which the evidence supports the various hypotheses you wish to consider.

The *Likelihood theory of evidence* claims that the only way in which evidence bears on the comparison of hypotheses (or models) is via the likelihoods of the hypotheses under consideration. Likelihood theory uses the Law of Likelihood (so named by Hacking, 1965) as a principle about the comparison of different hypotheses:

The observations O favor hypothesis H1 over H2 if and only if $Pr(O|H1) > Pr(O|H2)$.

And the degree to which O favors H1 over H2 is given by likelihood ratio:

$Pr(O|H1)/Pr(O|H2)$

We just described a situation in which we use likelihood in order to compare two different hypotheses on the same evidence. What does plausibility mean for likelihoodists? A problem about using likelihood in order to estimate the plausibility of a hypothesis is that confirmation works well when you are comparing two different hypotheses on the same evidence, and all you want to know is if that evidence favors one hypothesis over another. What you cannot know is if the same evidence could favor this hypothesis over other potential alternative hypotheses. In other words, from establishing that a hypothesis H1 has a higher likelihood than another alternative H2, it does not follow that H1 would be favored over other possible n alternative hypotheses H3,…, Hn. This is another face of the problem of the *not*H hypothesis, that shows how evaluating the plausibility of a hypothesis by itself in a likelihoodist framework does not work as well as comparing it with another.

Another limit of Likelihoodism can be that often you can consider completely absurd hypotheses that have a high likelihood. In this case, it is only your background knowledge which can tell you that you cannot accept an absurd hypothesis, also if it is highly favored by the evidence. So, also if likelihoods and the likelihoodist epistemology are considered more “objective” than priors, posteriors, and Bayesian epistemology, in a situation in which you have a paradoxical hypothesis with a high likelihood, it is only our subjective feeling or common sense that can tell us to do not take it in consideration.

We saw that the main difference between Bayesian and Likelihoodist theories of confirmation is the fact that Bayesians consider priors when they are going to evaluate the plausibility of a hypothesis. However, likelihoodists reject the idea of using priors. If we are considering simple hypotheses, inside both the two frameworks likelihood is well defined. But problems arise when you consider composite hypotheses and we have just seen that the likelihood of a model needs priors to be fixed in order to be defined. Now, while Bayesians give a value to these priors, likelihoodists do not and so in a likelihoodist framework the likelihood of a model is not well defined. A model does not have a well defined likelihood, but it does have a well defined likelihood function, because every hypothesis of the model has a well defined likelihood.

There is another way of associating a likelihood with a model. We saw before that this way is called the best-case strategy (see Sober, 2002a) and it consists in calculating the Maximum likelihood of the model. Instead of calculating an average of the likelihoods of the simple hypotheses of the model, you take the maximum likelihood of the model. What you obtain at the end is $P(O|L(M)) \geq P(O| Hi)$, for all Hi in M. The Maximum likelihood of a model (the likelihood of the likeliest special case) is always, by definition, greater or equal to the likelihood of the model, when the likelihood is well defined.

6.6 The Principle of Total Evidence

Bayesians and likelihoodists both agree on this principle. The principle of total evidence is considered one of the main principle of the theory of confirmation. It says that our beliefs on an inference should be based on the total evidence available at the time. Therefore, when we make an inference, we need to consider the total evidence we can have at the moment, if we want this inference to be epistemologically relevant. So far, this principle could be easily considered a silly principle by someone who is not familiar with it. For it seems to represents the story in the “Ficciones” by Luis Borges, about the picky cartographer who wants to draw a map of the world, and tries over and over again, looking for the most accurate map and at the end he draws a map which is the exact copy of the world! But, the principle of total evidence does not claim that we should take account of the whole information, but only the whole information available. I am going to clarify this concept trough its meaning in theory of probability, but before that I want to present it with Hempel’s words.

> Broadly speaking, we might say that according to this requirement, the credence which it is rational to give to a statement at a given time must be determined by the degree of confirmation, or the logical probability, which the statement possesses on the total evidence available at the time. Alternatively, that credence may be determined by reference to any part of the total evidence which gives to the statement the same support or probability as the total evidence: In this case, the omitted portion of the total evidence is said to be inductively irrelevant to the statement, relative to the evidence actually used (Hempel, 1965:64).

Let consider now an example in a Bayesian framework. The idea is that if you have an evidence E1 which confirms your hypothesis H over another *not*H and then you find an additional evidence E2, independent of the previous evidence E1, which also confirms your hypothesis H over *not*H, you can say that your hypothesis is more strongly confirmed. Two or more independent observations always discriminate between your hypothesis H and *not*H better than either of them could do by itself. This idea can be expressed in terms of likelihood like this:

$$\frac{\Pr(E_1 \ \& \ E_2|H)}{\Pr(E_1 \ \& \ E_2|\mathit{not}H)} > \frac{\Pr(E_i|H)}{\Pr(E_i|\mathit{not}H)} > 1, \qquad \text{for each } i = 1, 2.$$

where both E1 & E2 support H over *not*H

And this is because E1 and E2 are independent, that is:

$$\frac{\Pr(E_1 \ \& \ E_2|H)}{\Pr(E_1 \ \& \ E_2|\mathit{not}H)} = \frac{\Pr(E_1|H)}{\Pr(E_1|\mathit{not}H)} \times \frac{\Pr(E_2|H)}{\Pr(E_2|\mathit{not}H)}$$

And because both E1 & E2 support H over *not*H (each of the ratio is greater than 1).

The principle of total evidence does not tell you how you should describe the evidence that you are considering. You can describe it in different ways and this does not contradict the principle. This is an aspect that the cartographer should have considered. For example, I am playing cards and I tell you that I took four cards from the deck and three of them were kings. Then I ask you to calculate the probability that the next card is a king. You could be interested in knowing what kings were, but this information is totally irrelevant in respect of what I ask you. The same is when I toss a coin and I describe my data to you by saying that I had 6 heads in 30 tosses. Then I ask you to calculate the probability of getting a head on the next toss. Again, the information about the exact sequence in which I got the 6 heads is not relevant. A proposition in which I speak only about the sample frequency (in this case, 1/5) is logically weaker (in the sense that latter entails former but not conversely) than a description of data in which I tell you the exact sequence of heads. But still I can use the former without contradicting the principle of total evidence if I represents the strength of evidence by the likelihood ratio (see Sober, forthcoming). Let briefly see why. I consider two hypotheses about the probability of getting heads when the coin is tossed: p=1/4 and p=3/4. Now consider all the N exact sequences in which you get 6 heads in 30 tosses. Each of these exact sequences have a probability with a value of

$$N\ p^6(1-p)^{24}$$

Then, I want to know the likelihoods of these two hypotheses and I shall consider two different way of representing the evidence. In the first one I represent the evidence with the exact sequence of heads I have obtained and in the second I represent it only by the frequency of heads.

$$\frac{\text{Pr(exact sequence|p=1/4)}}{\text{Pr(exact sequence|p=3/4)}}=\frac{(1/4)^6(3/4)^{24}}{(3/4)^6(1/4)^{24}}=3^{18}$$

If we use this logically weaker description of data, we obtain:

$$\frac{\text{Pr(6heads|p=1/4)}}{\text{Pr(6heads|p=3/4)}}=\frac{N(1/4)^6(3/4)^{24}}{N(3/4)^6(1/4)^{24}}=3^{18}$$

The N’s have cancelled and you get the same result you have with considering the exact sequences. So, you do not need to use the stronger representation of data, since it makes no difference to the likelihood ratio. In this sense the sample frequency is a *sufficient* statistics.

6.7 Akaike Criterion and Model Selection Theory

Model selection theory is a part of statistics interested in providing methods for comparison of models, which began in statistics with a paper by Hirotugu Akaike in 1973.

We have seen that a model is a disjunction of simple hypotheses and also that models can contain some not fully realistic constraints or assumptions, which you can call *idealizations*. These assumptions can be false, but still the model can make accurate predictions, since the false assumptions do not drive the model too far from the truth. Obviously, speaking about "not fully" realistic assumptions and models to be "too far" from the truth are vague concepts, but I shall not analyze this problem here. What I would like to point out is the fact that, by using Akaike's idea of model selection, you avoid the problem of working with models which you know to be false, because Akaike's framework provides a way of thinking about model selection in terms of the goal of *predictive accuracy* and not truth. The predictive accuracy of a simple hypothesis can be defined as the average fit (measured by the log Likelihood) of the hypothesis to new data. The predictive accuracy of a model (composite hypothesis) is the average predictive accuracy of the Maximum Likelihood hypotheses of the model fitted to new data sets of the same size as the actual data set.

Akaike's criterion for comparing models does not tell you anything about which model is true, but instead it does tell you how high its predictive accuracy is. Because of the goal it analyzes, this criterion makes more sense inside an instrumentalist framework (see about this issue, Sober 2002b), than inside a realist one. For instrumentalism holds that the point of science is to provide accurate predictions and not to tell us which theories are true. Let me make three points about some relations between truth, predictive accuracy, and simplicity inside models:

- A true composite hypothesis can be less predictive accurate than a false one.
- If two models are nested the simpler entails the more complex.
- If the models are nested the more complex model has a higher maximum likelihood than the simpler model.

Where, given two models A and B, A is nested in B *iff* all the simple hypotheses in A are in B. If A is nested in B, A logically entails B. Consider the two nested models A and B (where *A* is nested in *B*):

$$A = h_1 \vee h_2 \qquad\qquad B = h_1 \vee h_2 \vee h_3 \vee h_4 \vee \ldots$$

$$\text{if } A \Rightarrow B \text{ then } P(A|O) \leq P(B|O) \; for\, all\, O$$

Philosophy of science has already considered the relations between truth and predictive accuracy, truth and simplicity. E. Nagel (1961), speaking about the difference between realism and instrumentalism, suggested that it was not substantive, in the sense that true theories make more accurate predictions than false ones do. Popper (1934) considers the relation between simplicity and falsifiability and argues that simpler theories cannot be more probable than complex if they are nested.

The Akaike framework considers the relation between simplicity and the fit to data in a model. Simplicity is expressed by the number of adjustable parameters in the model, *k*. The fit is measured by the logarithm of the maximum likelihood of the model. What you can estimate by using this framework is how accurate a model will predict new data when fitted to old. Akaike proved a theorem that bears on it and says that:

An unbiased estimate of the predictive accuracy of model $M = \log\left[\Pr\left(data|L(M)\right)\right] - k$

Note that because the value of the maximum likelihood (and consequently its logarithm) increases by making the model sufficiently complex, a penalty for complexity (–*k*) is introduced. This penalty makes AIC estimation unbiased.

You use the old data to find the likeliest member of the model *M* and then the natural logarithm of its likelihood. Finally you subtract *k*. From Akaike’s theorem it follows the model selection criterion:

The AIC score of a model *M*, AIC(*M*) $=_{\text{def}} \log\left[\Pr\left(data \middle| L(M)\right)\right] - k$

This criterion is used in a comparative way, in the sense that you are not interested in knowing the absolute value of a model's AIC score. Instead, you want to compare the scores of models that are fitted to the *same* data set. Note that the AIC scores of two models *depend* on the character of the data, in the sense that two models can have different scores with different data sets. It can happen that a model *M*1 has a better score than a model *M*2 with a data set A, but M2 scores better than M1 with a data set B (see below).

AIC is about a trade-off between fit to data and simplicity. It provides a principle to evaluate two aspects of a model: the variance and the bias. Consider a model *M* which is a set of curves $\theta_1 \vee \theta_2 \vee \ldots$, where θ^* is the true curve and $\hat{\theta}$ is the model fitted to a given data set *D*. The variance is connected with the idea of precision of a model and it represents the mean distance between θ^* and $\hat{\theta}$. The bias is the distance of the most predictively accurate hypothesis in the model *M* from the true hypothesis θ^*, which can be understood as what a model could do potentially. While *k* can be considered a way to measure the variance of the model, the logarithm of the maximum likelihood can be considered a way to estimate the bias of the model. Simplicity increases the bias of a model and decreases the variance of it. This is why complex models (with a large number of adjustable parameters) have large variance and small bias.

Akaike's theorem is based on three assumptions. One is that the process generating the distribution of the old data set is the same as the process generating the new data set (see Forster and Sober 1994). The second assumption is that the domain in which the old distribution of data falls is the same of that one in which the new distribution falls. Note that these two last assumptions constitute a problem of *Uniformity of nature*. In other words, it is a problem of *extrapolating* from a certain range of values to another, obtained from a different process. The problem is: Is an inference – reliable in a range of values - also reliable outside of that range? (see Forster 2000).

The third assumption is the normality assumption and it says that repeatedly estimated values of the parameters conform to a normal distribution.

There is a problem inside the AIC's comparison of two models and it concerns the size of the data set. Consider two models which are sets of curves and are nested. LIN ($y = a + bx + e$) is a disjunction of all straight lines in the X-Y plane and PAR ($y = a + bx + cx^2 + e$) is a disjunction of all the parabolas. LIN is nested inside PAR. The AIC's comparison of these two models changes as the size of the data increases. The fact is that when the amount of data increases, only the quantity of logarithm of likelihood will change, while *k* is a constant and so will not. We know that AIC(LIN) > AIC(PAR) precisely when

$$\log\left[\Pr\left(data \middle| L(LIN)\right)\right] - \log\left[\Pr\left(data \middle| L(PAR)\right)\right] > -1$$

where -1 represents the fact that PAR has one more adjustable parameter than LIN.

The equation can be rewritten as

$\log\left[\Pr\left(data \middle| L(LIN)\right) / \Pr\left(data \middle| L(PAR)\right)\right] > -1$, and (since the logarithm in base $e \approx 2.72$), this is true when

$$\frac{\Pr\left[data \middle| L(LIN)\right]}{\Pr\left[data \middle| L(PAR)\right]} > \frac{1}{2.72} \approx 0.37$$

This inequality, which describes the relation between the two AIC scores, may be true for small data sets but if you sufficiently increase the data set, it will be false. For large data set PAR will score better than LIN. The greater simplicity of LIN over PAR can compensate for L(LIN)’s lower likelihood for some sample sizes, but it cannot for sufficiently large ones. From the fact that LIN scores better than PAR on the old data set, it does not follow that LIN will score better also in a much larger one.

CHAPTER 2

THE HUMAN HISTORY: EPISTEMOLOGY AND METAPHYSICS

> *When I was very young, I was suitably impressed to learn that, appearances notwithstanding, the whale is not a fish. Nowadays these questions of classification move me less; and it does not worry me unduly when I am assured that history is not a science.*
>
> (Edward H. Carr, *What is History?*)

1. Probabilistic "Mode" and "Ideal Types"

In this chapter I am going to present some issues in the modern philosophy of history, which I consider epistemologically relevant for the analysis of particular aspects of phylogenetic inference. In particular, I shall highlight some epistemological aspects of the philosophy of M. Oakeshott - his idea of the "modes" of knowledge - and some aspects of the philosophy of M. Weber - his idea of the "ideal types". I am interested in their epistemology since I would like to argue that it is useful to understand the reconstruction of the phylogenetic history by means of probabilistic theories; for this reconstruction is characterized by a probabilistic "mode", with its epistemology and ontology of probability and every probabilistic model inside this mode can be thought of as an “ideal type”, in the sense of a framework of assumptions on the historical process, that makes possible a historical representation[1] of the evidence.

2. The Semantic Issue: What does ‘History’ Mean?

The term ‘history’ is ambiguous, because it is used with at least three different meanings. For history can be a) what happened in the past, b) the reconstruction of what happened, and c) the development of a project trough the past, present and future.

Usually, when by history we mean the reconstruction of what happened, we are focused on the object of study, the past, and on the method and epistemology of that reconstruction. However, when we accept the idea of history as a structure of past, present and future, we are focused on the *final meaning* of history.

I suggest a definition in which history is not the reconstruction of events that happened in the past, but the representation of evidence in a historical perspective. It is not the object which makes the difference, but our perspective. By accepting this definition, the focus is not on the nature of the object at all, but on the subject and the historical hypotheses used to understand and explain the evidence under a historical perspective.

The two definitions b) and c) are hold inside two different philosophies of history that are interested in this two aspects of history. The first one focuses on the epistemology and the method of the reconstruction of the past. I shall name this *analytic* philosophy of history. Analytic philosophy of history and historiography answer different questions about the knowledge of the past.

The second kind of philosophy, however, is interested in the development of history in the time, *in toto*, where history is interpreted as the confirmation of a plan on a higher

---

[1] See about the concept of representation in philosophy of science, Boniolo 2007.

level, like for example a theological plan. History is not the past but the totality of time, therefore, past, present and future. Past, present and future have their *sense* inside a historical project, which philosophy and history should understand. The past and the present, in particular, do not have a meaning in themselves and in their time, but inside a perspective that is available only in the future, when the project is fulfilled. The aim of historians must not be simply reconstructing the past, but understanding the *essence* of history and its final meaning, that develops on a non physical level and on a large temporal scale from the past trough the future. I shall call it *essentialistic* philosophy of history. In modern philosophy, this peculiar philosophical approach goes from Hegel and the German historicism to some modern forms of idealist philosophy of history.

This view of history, with all its metaphysical theories, has been for obvious reasons considered far from science and close to the idea of a *novel*. For the meaning of a novel like the meaning or in general a *sense* of history[2] is inside the novel or history itself. K. Löwith in *Meaning of History* defines this particular philosophy of history as "the systematic interpretation of universal history, where events and the historical developments are unified and go in the direction of an ultimate meaning". By this view, historiography and philosophy of history should aim to discover the universal laws that determine the human actions, which lead history to reach its final meaning.

In this sense of history any historical prediction in the future can be interpreted as a *prophecy*. A prophet talks about the future in a historical perspective, as the possible carrying out of a historical well defined plan. The present reaches its meaning in the future, when it has become past and it becomes part of the development of history[3]. We shall see that Popper gives to this kind of philosophy of history the label of "historicism".

---

[2] I use the term 'sense' of history in a very broad range of significances. By 'meaning', instead, I refer to the hermeneutic tradition in the modern philosophy of history.

[3] See about this idea of history Löwith 1949. See about the idea of prophecies in history, Danto 1965.

While there are three main definitions, there are two possible ontological positions about the past:

- That one which has metaphysical implications, by focusing on the nature of the past.
- That one which does not have metaphysical implications and focus on the role of representations in our historical understanding, independently of the nature of the past.

Definition c) has metaphysical implications; definition b) can have some metaphysical implications, depending on how we use the term 'reconstruction'; my definition, however, does not.

Finally, we shall see that there are three possible epistemologies of history: the epistemology of interpretation of thoughts; explanation of events; representation of evidence.

3. History: Explanation of Events vs Interpretation of Thoughts

> The historian, investigating any event in the past, makes a distinction between what may be called the outside and the inside of an event. By the outside of the event I mean everything belonging to it which can be described in terms of bodies and their movements […] By the inside of the event I mean that in it which can only be described in terms of thought […] The historian is never concerned with either of these to the exclusion of the other. He is investigating not mere events (where by a mere event I mean one which has only an outside and no inside) but actions, and an action is the unity of the outside and inside of an event. (Collingwood 1961, p. 213)

I have said that the definition of history as a reconstruction of the past focuses on the object and the knowledge of that object. However, philosophers and historians have had different opinions about what "past" should have been considered.

While inside positivism history was considered as a reconstruction of *facts*, inside the idealist philosophy of history of B. Croce, M. Oakeshott and R. Collingwood, history was the history of human *thoughts* leading the actions. The historian can understand history only by understanding people's thoughts and their values. This idea of history mainly comes from the German tradition of Dilthey's historicism, in which the study of history was "interpreted" as an *empathic comprehension* of people's experiences.

The idea of the existence of two faces of history, the inside (thoughts) and outside (actions), has been recently translated in the idea of the *structure* and *system* of history by the historian and archaeologist P. Kosso (2001). The idea is that archaeologist should be able to distinguish a "system" and a "structure". The former includes the material things and their physical organization: the buildings and their arrangement in settlements and towns and interactive networks. The structure, instead, includes the values of a society and the ideas and intentions of individuals of that society. The system is the manifestation of the structure. Kosso individualizes the importance of an approach to history *from inside*, which is exactly the idea of *history of thought* we find in Collingwood.

> The process of nature can therefore be properly described as sequences of mere events, but those of history cannot. They are not processes of mere events but processes of actions, which have an inner side, consisting of processes of thought; and what the historian is looking for is these processes of thought. All history is the history of thought. (*Ivi*, p. 215)

The interpretation of the historical object as the processes of thought behind the human actions had as a consequence the idea that history and science should have been considered separated and studied with a different method and epistemology.

One of the idealist philosopher of history who hold this dichotomy between history and science, even if in an original way, was Oakeshott. Oakeshott in *Experience and its modes* expressed this idea by dividing the experiences in different "modes"; therefore there is the mode of history and the mode of science. For Oakeshott experience is organized in different modes. The modes are worlds of ideas, thoughts, where each world is auto-consistent and coherent. Experience exists and it can be true or false only inside one of the mode with all the presuppositions and concepts we have considered inside it. Since the modes give a meaning to the experience which does not exist independently from them, the world of the experience is the world of ideas. These worlds or modes represent different points of view of the same whole experience, they are coherent in themselves but not complete; they are *abstractions* of the whole experience, independent from each other and incommensurable. Nevertheless, even if the modes are independent from each other, they are dependent of the totality of experience, insofar they are forms of the experience. A mode is "the whole experience arrested at a certain point and at that point creating a homogeneous world of ideas"; "I mean, then, by a mode of experience homogeneous world of ideas" (Oakeshott 1933, pp. 74-75). Experience is not reducible to the summary of the all modes and there is no possible knowledge in the sense of a *correspondence* between modes and experience, since there is no experience outside of a mode. Knowledge is possible only inside the modes and for this reason it is relative to the mode we choose and determined by its coherence. The experience acquires an ontological – in the sense that becomes real - and epistemological status - in the sense that is believed to be true or false - only inside one of the modes; outside of it, experience is meaningless, without any ontological or epistemological status. A mode of experience, insofar as it is coherent, is true for itself, but it is never necessary to experience and unavoidable.

> And the character of this world of ideas, since it is created by abstraction, is to be determined only by considering its coherence, that is, by considering how far it succeeds and how far it fails to provide what is completely satisfactory in experience. […] No abstract world of ideas is independent of experience, for each is experience at a certain point. And no abstract world of ideas is independent of the totality of experience, for each derives its character from the whole from which it is an abstraction, and apart from that whole it has neither meaning nor significance. (*Ibidem*)

Oakeshott's philosophy represents a peculiar version of idealism, insofar the abstracts worlds of ideas are not independent of the experience. It seems that the relationship between reality and worlds of ideas is not one way, like it is in idealism. The world of experience has no meaning outside of the world of ideas, but the same it is true for the world of ideas. A mode has no meaning apart from the experience from which it is an abstraction. Moreover, modes are not permanent forms arranged, as the Idealists were inclined to argue, in a necessary and hierarchical manner; they are historical, mutable creations. Each mode brings to experience its own criteria of reality and truth.

What is for Oakeshott the historical inquiry? It is a construction of the past from the explanation of the evidence. To explain an event historically is to relate it to antecedent events so that its character can be understood by this relationship. Oakeshott believes in a distinction between scientific and historical mode and he thinks that only history can provide the model of explanation in the human sciences. Therefore the historical construction is made by recognition of events as the outcome of events to which they are contingently related and trough the understanding of the human conduct which consists in choices and actions. However, even if there is a methodological distinction between science and history, there is not a dichotomy on a metaphysical level. They are two modes of the same knowledge, and while history is not a description of the human world as it is, so is science. Like history, science organizes experience according to its own categories and presuppositions, including its own criteria of reality and truth. They are distinct in the sense that they have different ways of understanding the reality, a different ontology and epistemology. For Oakeshott human understandings have their histories but these histories are contingent in the time, in the sense that there is not a progressive enlightenment of the knowledge. There are not in his philosophy metaphysical elements like the idea of history as the story of the teleological development of “mind” and “reason”, like in the Hegelian idealism, for example. In the sense in which he considers knowledge as concepts-mediated but contingent, conditional and lacking of metaphysical elements, Oakeshott is certainly closer to a neo-Kantian philosopher like E. Cassirer (1910). In another sense, he is close to the hermeneutic tradition, since for him experience is made by meanings and therefore any knowledge with its concepts is interpretative. However, he does not bring to the extreme consequences the hermeneutical approach by saying that human and natural science are undistinguished because of their interpretive nature. Even if experience is made by meanings, modes rest separate ways of interpretation. The historical mode is different from the scientific mode because it has to interpret two levels of meanings instead of one. The levels of meanings we give to the reality in order to grasp it and the level of meanings that are within the historical reality, that is within the actions of humans. Therefore the historical mode involves an interpretation of the interpretation. The difference is epistemological, but not metaphysical, because it does not lie in the nature of objects we want to understand, but in how we should categorize them.

4. Continuity in Historiography

Historiography is characterized by different kinds of *continuum*, connected to each other.

I shall consider here the continuum between past and present and the continuum between history and its reconstruction.

In Giambattista Vico's *La scienza nuova*, for example, we can find a *metaphysical interpretation* of the first continuum. History can be thought of as having some patterns that repeat themselves over and over again in the time. In this perspective, there are a certain number of possible structures in which the particulars can change in different temporal and spatial context, but still the pattern to which these particulars events belong is the same. If one believes in this possibility, present and past represent different scenarios or moments of the same historical pattern. From this idea of *uniformity* in historical patterns follows the possibility of making prediction in the future by studying the past patterns. However, one can also interpret these patterns in a non-metaphysical sense, for example like methodological tools used to understand history, like common conceptual frameworks used to establish analogies between past and present.

What about the epistemological aspects of the continuum?

History is sometimes considered peculiar since it is a reality that cannot be observed. Is the non physical accessibility of the past a problem in epistemology? My answer is negative and in order to justify it I would like to use an argument from philosophy of science. For there is an analogy of this continuum between past and present with the continuum between observable and non-observable entities, based on the fact that of both past and non-observable entities we can observe only the traces they leave, which constitute the evidence of the phenomenon we want to explain. If we do not accept a difference between the knowledge of observable and non-observable objects, there is no epistemological reasons for considering history peculiar only because the past is not physically accessible. The dichotomy is not between past and present, but between historical and non-historical representations of the same evidence. Inside the essentialistic and idealist philosophy of history this idea of continuum is certainly accepted. In the first one, because history is in a continuum development from the past to the future; in the second one, since history is the historian's interpretation of past thoughts (see below).

In Oakeshott, history is a *world of thought*, it is the continuous affirming of a past that is not a past and a present that is not present. History is always in the present since, even if the facts seem to be so far in the time, history always refers to the present needs and situations and reflect the values of the present societies. In M. Bloch (1953) we find the idea that the comparison between events in the historian's craft requires a space-time simultaneity. D. Lowenthal in his *The past is a foreign country* (1985, p. 28-34) imagines the past as a foreign country where the historian travels, breaking the space-time gap with the present. We can also consider history as the "creation of a continuity". E. Hobsbawm and T. Ranger (1983), for example, claim that traditions are invented with the aim of transmitting values and norms of behaviour by repetition, which implies continuity with a suitable historical past.

Another continuum is between history and its reconstruction. In philosophy of science this would be the problem of *theory ladeness*, the issue regarding the possibility of an "objective" knowledge and observation, independent of the subject-observer. Is it possible to reach an objective history from the analysis of the sources? How much of the historian's interests and points of view is in our historical reconstructions?

In an idealist framework history is the historian's experience of the thoughts of people of the past and in this peculiar sense it is subjective; this subjectivity, however, does not mean necessarily that history must be considered a construction without any form of objectivity. History in this framework is subjective in the sense that it is an interpretation of thoughts instead of being a reconstruction of events. But from this it does not follow that it is not possible to believe in a univocal history as an univocal interpretation of thoughts and consider it in this sense objective. In this epistemological framework the reconstruction of history can be interpreted as a continuum between the historian's thoughts and the thoughts of people from the past.

Consider now history as a discipline explaining the evidence under a historical perspective, therefore outside of the idealist interpretation of it; in this case we should consider a dichotomy between construction of events of history vs reconstruction of events. The term 'construction' of events stresses the total subjectivity of our historical theories, in the sense that they can not be justified independently of us; it refers to an idea of history as a creation of our mind, where it is explicitly claimed that the past does not exist as ontological entity separated by its constructions.

The term 'reconstruction', however, refers to a naïve realist idea of history as a collection of facts and historical theories as "mirror-images" of the causal relations among past events. The dichotomy should dissolve in a third position, in which history is nor a construction, neither a mirror of the past, but it is the historical representation of the present, independently of any metaphysical claim. Therefore by 'historical representation' I mean a non-naïve reconstruction of the processes that can explain the present evidence, through a structure of models, hypotheses and historical interpretations joint together.

We have seen that Oakeshott does not believe in a naïve reconstruction of the experience in the sense of being non-mediated by interpretative concepts, even if he does not say that there is no reality independent of our knowledge of it. He does not claim that history is a mere construction, since each mode is not independent of the experience, but that there is no knowledge of the reality independent of our ideas. He does not consider history as a science, since they are two distinct modes of experiences, and in this sense he is still close to the German idealism of Dilthey, but he interprets the reconstruction of history inside a Kantian epistemology, and in this he is far from idealism and close to the E. Cassirer's philosophy of science.

> History, then, begins not with the collection of isolated particles of data, not with a universal doubt, nor with a blank and empty consciousness, but with a homogeneous world of ideas. […] I take it, then, that *history is experience, and not a course of events independent of experience*. History is not the correspondence of an idea with an event, for there is no event which is not an idea. History is the historian's experience. (*Ivi*, pp. 98-99. Italics added)

We can compare this claim with the following Cassirer's claim and see how close they are.

> Thus no content of experience can ever appear as something absolutely strange; for even in making it a content of our thought, in setting it in spatial and temporal relations with other contents, we have thereby impressed it with the seal of our universal concepts of connection, in particular those of mathematical relations. The material of perception is

> not merely subsequently moulded into some conceptual form; but the thought of this form constitutes the necessary presupposition of being able to predicate any character of the matter itself, indeed, of being able to assert any concrete determination and predicates of it (Cassirer 1910, p.150).

There is no reconstruction in the sense of a naïve realist position in which the mind is a mirror of the reality, but it is clear that for both Oakeshott and Cassirer, even if they are considering different "modes", respectively the historical and the scientific one, experience is already conceptually laden and therefore distinct from the reality itself.

Another position that supports the idea of a non-naïve reconstruction of history is that one offered by E. Carr, who defines history as a dialogue, an interaction, a correspondence between the historian and the facts he/she decides to consider in a historical representation. By paraphrasing the famous Kantian image of reason, he claims that the historian without facts is rootless and futile, as facts without the historian are dead and meaningless.

> The historian and the facts of history are necessary to one another. The historian without his facts is rootless and futile; the facts without their historian are dead and meaningless. My first answer therefore to the question, What is history?, is that it is a continuous process of interaction between the historian and his facts, an unending dialogue between the present and the past. (Carr 1961 p. 24)

History can be seen as a mere mental construction. In this view events do not exist independently of the history that historians write about them. We find this idea in the extreme position of the American historian C. Becker (1910) claimed that historical facts do not exist until historians create them.

Let us see now some interpretations of history as art. What does it mean for a history to be as a art? The most famous of these interpretations of history, is the one of history as a novel, which has dominated much of romantic historiography. However, in this view of history another aspect of subjectivity is highlighted; it is not the idea of the conceptual framework that we use to make inference about the meanings of the past events, but it is more the idea of the selection of the evidence that historians make in their reconstructions. History can be represented by a painting. It is possible to create some ideal representations, where events from a historical period are compared with other events from a different period. For J. Gaddis (2002) history should be *represented* more than *reconstructed*, and any representation is a landscape of a choice of non-contemporaneous events. Historians are painters, and what they represent is the product of a certain choice and point of view. Consequently, any representation has a specific meaning, by reflecting the aim of the historian-painter and the interests of the society and the time in which the historian lives. The subjective selection of evidence that historians make in their historical representations is necessary for the meaning of the representations, since they are the answer of our questions. Moreover, historians should add to the sources their own sources, coming from both their *logic* and *imagination.* This is the idea of the *thought experiment*, as used in J. Diamond's *Guns, Germs, and Steel* (1997) in order to give a possible explanation to the contemporary inequality through-out the world among different societies.

For, given that all histories should be thought to be conceptually laden, the comprehension of the presuppositions and perspectives will make possible to justify different histories and accept them as contingently and conditionally objective. *An* objective history does not necessarily mean *one* objective history. We can have more representations of the same history, answering different questions in different contexts.

The contraposition that the German historicism makes between the view of a history characterized by general patterns and a history characterized by the infinite experiences of each single individual, disappears in the image of history as a painting, where one can represent both the general and the particular, only by focusing on one instead of the other. It depends on what we are interested in, what we consider to be problematic, but all the images reproduce the same object accurately. Nevertheless, relativity and pluralism is not an essential aspect of the nature of history, as many times has been claimed. It is not the peculiar inaccessibility of history which forces us to accept a pluralistic approaches. We choose a pluralistic way of knowledge since we think that it is a way of reaching more accurate and predictive representations of it. I prefer to use the term ‘conventionalist’ approach as it is used by Poincaré in *La science et l’hypothèse*, as I have introduced it in chapter 1 about the problem of species. The idea is that different historians analyse different aspects of the same facts, starting from different social and historical points of views. The different resulting histories have the same explanatory value, they cannot considered in competition to each other, since they are answering different questions.

In the knowledge of the past there is not only the possibility of choosing many points of view, but also many different facts. Often the past in the literature has been defined as a *mish-mash* of facts. In this continuum of facts historians, like detectives, are used to choose some clue-events, in order to give a meaning to history. After selecting particular clue-events, historians also select a particular framework of time and space and inside this framework they put all events. No event has a value in itself, but only inside a framework and in relation with the clue-events. The final version of the framework is the reconstruction of the history. We have seen above that Gaddis uses the metaphor of a painting, where reconstruction of history is a landscape in which historians put the events together in a common space and time. The idea is that the epistemology of history is based on a simultaneity of space and time, because the understanding of any event needs the comparison with other events in a common framework of space and time. His paintings or landscapes of history are analogue to the Weber's ideal types and Popper's interpretations (see below).

It follows a table that resumes the main positions about ontology and epistemology of history.

Ontological issue: What is history?	Epistemological issue: How do we know history?
Naïve reconstruction of the past. Metaphysical implications	Explaining events
Mental construction of the past. Metaphysical implications	Interpreting ideas
Process developing from the past through the future	Understanding the final meaning
Non-naïve reconstruction of the past. No Metaphysical implications	Representing the evidence in a historical sense

5. The *Sense* of History

We have seen that history can be analyzed analytically by focusing on the reconstruction, construction or interpretation of the evidence and in this the problem of the knowledge of history are the same of the issues in philosophy of science, since they are just two ways of explaining the reality. However, there are some specific questions about history which are not about its method or its epistemology and they are addressed in particular by the idealism and what I have called essentialistic philosophy of history. These questions are: What is the "nature" of history? Is there a sense of history? What is this sense? Has history an aim? The answer of these questions have generated many metaphysical theories that I am not going to analyse here; I would like just to present roughly the interpretation of history as a process. Any experience can be thought as a unique experience or considered to be part of a historical process. From a metaphysical point of view this process can be considered either happening *out of* the time or developing through the *whole* time (past, present, future). This idea of history as a process independent of the time represents the issue of the static theory of time in philosophy of time[4]. Inside this view of time, the reality has an a-temporal character where the future is well determined and real, exactly like the past. The statements about events are formulated in a way - defined as 'a-tensional' - that they can be judged to be true independently of the time in which they are judged. However, this epistemological possibility does not imply the existence of a-temporal or omni-temporal events.

In philosophy of history the historical process has been usually considered as the structure of all single human experiences developing in the whole time, where any single experience has a specific significance. For Croce (1917, p.190), history is made by all the individuals with their experiences. Any individual, even if considered meaningless, intervenes in some way in making history. In the movie "It is a wonderful life" (1946) by F. Capra an individual's life is considered as a necessary part of one of the possible historical chains of events. Any event is considered joint to another in a network of interactions that represents one of the possible stories where different interactions could have been possible.

In the movie "Judgment at Nuremberg" (1961) by S. Kramer, Nazis are judged for the evidence (the murder of people) under a non-historical perspective. Their actions are judged under a non-historical ethical perspective. In the present they are judged as individuals who have killed other individuals. But the Judge, at the end of the process, claims that it is not possible to judge this people, for the same evidence, under a historical perspective. That judgement will take place in the future. Insofar their actions

---

[4] See for an introduction to the philosophy of time, e.g. Dorato 1997.

are part of the historical process, they must wait the "end" of history in order to be judged in a historical perspective. The idea is that the meaning for their actions in the historical perspective should be considered inside the whole history. History in this sense is not the reconstruction of events but the judgement of events.

We have seen the meaning of the two continuum between construction and reconstruction and past and present are correlated in the idealist philosophy. In the essentialistic philosophy, history is considered the interpretation of a meaning that develops in the time. This means that any event could be re-described and re-interpreted over and over again in the historical time, and its meaning could change in time. In this sense Popper is right when in *Poverty of historicism* (1944-1945) denies historical predictions. By accepting this particular metaphysics of history we should say that historical concepts and theories are non-projectible, like N. Goodman's "grue" and "bleen" emeralds[5]. The historical meaning of any event should be re-considered in the light of the information historian have in a certain time *t*, which can be different to the information they have at the time *t*-1 and *t*+1, depending on the relations it can have with future events. The historical meaning of something depends on the context one is considering.

We have briefly seen that history can be considered simply a reconstruction of something and therefore it can be an interpretation of thoughts, of values and intentions, an explanation of events, a representation of the evidence. But it also can concern metaphysical theories about the development of general patterns and meanings.

All these possible interpretations are correlated with different views linked to different historical periods. L. Ranke is credited with the idea of founding of scientific method in history. In his view, history was the science of telling things as they actually

---

[5] See Goodman 1954.

occurred in the past. History must be based on the documents, since if it is not in the documents, it does not exist. His position gave the birth to one of the several notions of history, in which the past has to be understood only through a careful analysis of sources. Positivism was closely associated with Rankean method. For in the positivist view there was the idea that science and its method could have help the progress of the knowledge of the history and the society. This view was considered by other thinkers to imply a reductive idea of history, as a sequences of events. Thinkers like Croce and Collingwood strongly opposed to the idea that historical events could be explained as other kind of events. In this view the scientific method cannot be applied to historiography because of the peculiarity of its object. Explanation in history cannot consist in the mere causal correlations of events. For history is made by humans and historical events hide thoughts and motivations of humans in the time. In this view the term *explanation* results to be not really appropriate in history, and other terms like *comprehension* or *interpretation* have been often chosen instead of it. The term interpretation is the same that Popper uses for the historical knowledge. He denies the possibility of considering history a science, because of the fact that history cannot create new general laws but it can only use “trivial” laws (see below) in order to interpret single events.

6. Historiography, Historicism, and the Analytic Philosophy of History

I should define more precisely some terms I have already introduced. By ‘historiography’ I mean the science of studying history. I shall refer later to two different approaches in studying history: the analytic philosophy of history, which I have defined above, and the *historicism*. By ‘analytic philosophy’ of history I mean the “critic” branch of philosophy of history, developed in the 50’s, that analyses the methodology and epistemology of history through the analysis of the language and focusing on the logic of historical explanation. This branch of philosophy was against any metaphysics, in particular the metaphysical distinction between science and history on the basis of the nature of their object. 'Historicism' is an ambiguous term that is used in philosophy of history and history itself, reflecting a specific metaphysical and "speculative" approach to the study of history. A 'speculative' approach to philosophy of history refers to an idea of history as a set of single human events, unique inside their own society and time. In this concept of history, the historical change is defined through the diversity of values that characterizes any society in any time. In 1800 this idea of history represented one of the reaction to the empiricist and instrumentalist approach of positivism, which aimed to reduce any discipline to science for the progress of society and humankind. In general, in the literature, the term 'historicism' refers to that idealist philosophy of history that considers history a human science separated by the natural science, because of the peculiar "nature" of human sciences. In particular, the term historicism is associated to the German historicism. In Popper's *The Poverty of Historicism* there is a particular definition of 'historicism'. Here 'historicism' is the philosophy that interprets history in a deterministic view, therefore with all that metaphysics characteristic of the essentialistic philosophy of history. A deterministic view is founded on the existence of a sense of history, which has to be discovered in some kind of destiny or ineluctable forces. These forces are the ineluctable historical laws that historians aim to catch and use for predicting events. Popper recognizes the main representation of this kind of philosophy in Marx and defines his theory as a "historical prophecy".

> […] I mean by 'historicism' an approach to the social sciences which assumes that *historical prediction* is their principal aim, and which assumes that this aim is attainable by discovering the 'rhythms' or the 'patterns', the 'laws' or the trends that underlie the evolution of history. (Popper 1945, p.3)

6.1 Weber and the "Ideal Types"

M. Weber belongs to the tradition of the German historicism, but he proposes a new epistemological perspective in the field. Before Weber, a clear cut between the epistemology of the natural and human sciences was accepted, on the basis either of a strict metaphysical distinction between the natural and the human world[6], or on the basis of a distinction of the method to be used in the historical and natural sciences[7].

In his famous book on the methodology of social sciences, Weber focuses on the epistemological status of the human sciences, called also social, historical, or sciences of spirit. Weber adopts the distinction between generalizing and individualizing method, but he refuses to accept any kind of distinction between human and natural sciences based on the object and on the method. For him the historical knowledge is a science of the reality, since there is only one *science* and only one *scientific method.* Any science can be generalizing and individualizing, depending on the necessity and the aim of the research. By starting with this idea, for him the epistemological analysis of history is not focused on tracing the boundaries between different sciences but it becomes an analysis of the relation between concept and reality, law and history. For, none of the

---

[6] W. Dilthey (1883) is one of the modern philosophers of history who founded his epistemology of history on this metaphysical distinction. While the natural world is objective, the human world is the world of the historical and social relations among individuals. The natural world is the object of an external experience and the human world is the object of an internal experience.

[7] W. Windelband (1883) proposed the methodological dichotomy between nomothetic or generalizing and idiographic or individualizing disciplines. While the first ones try to individualize a system of general laws expressing regularities among phenomena and the second ones aim to analyze specific phenomena in their individuality. The nomothetic sciences are called natural sciences and the idiographic sciences constitute the historical sciences. While history explains single events, natural sciences explain general laws.

two methods is able to catch the whole reality. The knowledge we can reach with both methods cannot be the reproduction of the reality. By using one instead of another, only some aspect will be detected. So, depending on what we aim to know, we are going to choose one or the other. Weber was against the theory of empathy, the theory that history should be based on the *intuition*, the sympathetic *re-living* of experiences of other individuals. His idea was that *repeated experiences* of the others *cannot replace concepts*. For Weber there is only one science of the reality, and only one scientific method. In the causal explanation of phenomena, we always have to choose some events and some aspects of them, and the criterion is by referring to values. However, these values are not absolute and they are important just at the beginning of our process of knowledge, in the context of discovery, since they promote our interest and curiosity. They give a direction to our research by leading the choice of phenomena to be considered and by determining the questions to be asked about them. Phenomena are not important by themselves, but only by referring to a certain kind of political, cultural, ethical, religious values. It is on the basis of specific values that we can select events and aspects of these events, and the choice of these values is necessarily founded on our *points of view*. A phenomenon is “cultural” if and only if it interfere directly or indirectly with our cultural interests. Our knowledge is guided by values, that give a meaning to the part of the reality we are considering under our concepts. Our knowledge of history *is not a reconstruction* of an event in the sense that it can reproduce the whole event. After we have made a selection of phenomena, then our research continues with a *causal explanation*. History must not reflect events, but its aim is to give a causal explanation of those elements and aspects of events that have a specific historical meaning under specific values. The change of values changes the way in which we are considering the reality, since it changes the perspectives under which we see it. Weber considers absurd the idea that it is possible to reach a *final* system of concepts in which the whole historical reality can be caught and understood, since history is something in a continuous change. The change is not metaphysical, but epistemological. What changes are our conceptual structures and consequently the meaning we give to history. Therefore, a final form of the system of historical sciences represents a non-sense.

One of the most interesting Weber's theories is the theory of the "ideal types".

The "ideal type" is obtained in historiography when historians decide to select *one or some specific points of view* and on the basis of them they make correlations and create connections among different phenomena, with the construction of a unique *conceptual picture or illustration*. This picture of the reality cannot be found inside the reality itself, since it is just *utopia*. Each conceptual picture is an ideal type. It is not the historical reality, but it represents it in an ideal framework: it is an *example* of it. Then historians have to make a correlation between the reality and the ideal type, because just like this the reality becomes meaningful for us. The point we should stress is that the ideal type has not a metaphysical meaning. It is only a methodological tool used with a heuristic aim. The key idea is that the ideal type not only does not reflect the reality, but also it must be "far" from it, because just like this it can be used in order to catch the meaning of the phenomenon. The historical explanation is for Weber a causal explanation, where causes like the object of the research have not all the same meaning for any historian and they must be chosen. In order to understand the causal relations between events, historians need an *unreal construction*, a *hypothetical process* in which only some elements (causes) of the empirical content of the reality have been isolated. Then the historian compare the hypothetical process with the real one. The construction deviates from the reality as much as the elements which have been excluded in the hypothesis were important in the real process. In other words, the historian constructs with the evidence different *possible* scenarios, showing what it would have happen in the history if some events in the past had been different. By inventing a different possible outcome, the historian can show how important are some causes in the happening of the present. The fact that with another possible scenario, without one of the causes that are considered important for the happening of one present instead of another, we could have another present, it is an *objective possibility*. This possibility has an important meaning, because it highlights the non-deterministic view of history. The fact that a scenario of the past (cause) could bring to the present (effect) is *only sufficient* and not *necessary*. Different causal scenarios, characterized by different interpretative hypotheses, can be considered possible causes of our present, but with different probability. Moreover, given another possible cause we wish to consider in our construction, any effect is just equally probable. In a deterministic view of history just one causal-effect scenario would have probability equal 1, while the others would have a probability equal 0. A causal construction with a high probability is for Weber objectively possible and so can be considered an *adequate causation*. When its probability is very low, this construction must be considered an *accidental causation*. In other words, if the hypothetical process does not lead us to the phenomenon we aim to explain, the cause that has been excluded in it is linked to the real process with an adequate causation. However, if the constructed process produces the same effect like the real one, then the excluded cause is linked to the real process with an accidental causation. Weber does not want to give to the history and its development an immanent rationality, in the sense that he rejects a deterministic view of history. The only sense, direction and rationality is that one we give to it with our ideal types and causal constructions, which have just a certain degree of probability, therefore they represent just different possibilities. Historiography is for Weber a science, since scientific explanations of the past are possible and with them it is possible to reach *objective* pictures - characterized by cause-effect relations - with a certain degree of probability. The ‘objectivity’ is provided by the possibility of giving an epistemological value to those pictures of history, but it does not mean that it exists a univocal reconstruction of history.

6.2 The Anglo-Saxon Debate in Philosophy of History

The Anglo-Saxon debate on historical knowledge and method can be seen as articulate in the two main approaches I have introduced above. The speculative approach and that one of the analytic or critic philosophy of history. I shall analyse the debate inside the second approach, by focusing on the issue of the relation between history and science and the issue of historical explanation.

We have seen before that in the first German historicism of Dilthey, the background philosophy was the positivism and the idea of a reduction of the historical method to the scientific one. Later, with the Anglo-Saxon idealist tradition the background philosophy was the neopositivism and its battle against any metaphysics and a non scientific knowledge based on non verifiable statements. Inside the previous tradition of philosophy of history and in general human disciplines, the knowledge was charged of metaphysical values and not susceptible of empirical verification. Therefore, inside the neopositivistic view, there was only one way of accepting the human disciplines: to change them according to the scientific criteria. Neurath (1934) refused the ontological distinction inside the historicism between sciences of spirit and sciences of nature, based on the object and on the epistemological disposition of the subject towards it. He tried an unification of all sciences on the basis of linguistics. Neurath proposed a solution of the problem of knowledge inside the human disciplines in the psychology of behaviourism. He tried to empirically found the human sciences inside the inductive and pragmatic framework he adopted for natural sciences, by using the theory of behaviourism. Inside the framework of behaviourism the object of human sciences could have been considered a set of observable behaviours expressed as empirically testable propositions. He had the idea of a unique conception of knowledge and tried to unify all sciences under a common linguistic basis and the idea of a verification based on the correspondence between each proposition of observable events (and behaviours) and the whole complex of propositions that constitute the scientific knowledge. The scientific model to which any science should have tried to refer was physics with its laws and the idea of predictions testable in a spatio-temporal context. Sociology was for him the science of social behaviourism, with the aim of describing observational correlations between human behaviours and predicting testable observable propositions. He considered historiography a sort of empirical sociology. His idea was that all sciences should have been considered connected to each other, by means of the same scientific method and epistemology, characterized by: the cult of facts, a conventionalist approach in selecting evidence, and finally a pragmatic view about the meaning of predictions. However, Neurath, like Popper (see below), recognized the limit of social sciences in making predictions and using causal explanations where the generalizations play a double role: the role of classifying events and initial conditions; the role of using common patterns, regularities or laws in order to establish a relation between initial conditions and evidence.

7. History and Science

Is history a science? Two are the main points of view: that one of those who believe in the unification of the scientific method and that one of those who believe in a dichotomy between these two disciplines, based on a specific idea of history. Popper, however, develops a third position, with a new kind of dichotomy between historical and generalizing sciences. The unification of the scientific method accounts for the model of explanation and for the idea that general laws are logically necessary inside explanation. Popper believes in the unification of the model of explanation, but he refuses the possibility of general laws in history. While Hempel (1962) is convinced that historical laws exist, even if historians do not explicit them, Popper (1945) claims that historians *use* laws from other disciplines, but general laws cannot be discovered inside history. He calls the historical laws "prophecies" – a term that usually recalls a supernatural plan – by referring in particular to the atheistic and pragmatic Marx's philosophy of history, which aims to discover laws for predicting and manipulating the future history. Historical reconstructions, exactly like non-historical reconstructions, are connected with the terms predictions and explanation. In the following paragraph I am going to present the Deductive-Nomological (DN) model of explanation and the concept of causality in history.

7.1 Explanation and Prediction

"The main function of general laws in the natural sciences is to connect events in patterns which are usually referred to as *explanation* and *prediction*" (Hempel, 1942, p.35).

It is well known that in the model of DN explanation, also called Popper-Oppenheim model[8] and by Dray the "covering laws" model, explanation and prediction are just different aspects of the same structure. This is because while the model is one, what we consider as a issue can be different. In other words, the difference between explanation

---

[8] See about the deductive-nomological explanation Hempel and Oppenheim 1948.

and prediction is only a pragmatic difference. If the issue is represented by the initial conditions that, on the basis of the law we are considering, are sufficient to explain an event, we are interested in an explanation of that event. In this case the event is already known to be happened. Instead, whether the issue is represented by an event that can be expected on the basis of the law and the initial conditions, we are interested in the prediction of that event, which in this case is not already known to be happened. In his *Reasons and Covering Laws in Historical Explanation*, Hempel considers two kinds of explanations with the same model: the DN and the Probabilistic-Statistical explanation. The difference is only in the general laws inside the model. They can be deterministic or probabilistic. In the former case, we have a deduction of the *explanandum* from the *explanans*; therefore the explanans logically entails the expalnandum. In the latter case, the explanandum follows from the explanans only with a certain degree of probability. The historical explanations have often for Hempel a nomological character, in the sense that they aim to demonstrate that the event-explanandum follows from some initial conditions; in order to discover these conditions historians use some generalizations which have a probabilistic character.

Prediction, exactly like explanation, can be deductive and probabilistic. In the former case it consists in deriving a future event from specific initial conditions and suitable general laws. The probabilistic prediction consists in deriving with a certain probability the same event, which in this case it is not to be expected, given the law and the initial conditions, with a probability equal 1.

What about history? What is the meaning of explanation and prediction in history and can this model be applied in historical explanation and prediction?

First I want to show that while the term ‘explanation’ is not ambiguous in history, the term ‘prediction’ can be used with two different meanings. There are predictions in the past and predictions in the future. While the former kind of predictions is just part of a historical reconstruction, the latter is a prediction we make for the future history on the basis of some patterns we have observed in the past and present history. When Popper defines historicism trough the idea of historical prediction, he has in mind the possibility of applying general laws reflecting regularities in history, in order to make predictions in the future. He is against the idea that historians can find any uniformity in history. The meaning of a prediction is immediate for the future, but not for the past, since in the past anything we can predict has already happened, like for scientific predictions of the origin and evolution of organisms. Making predictions in the past represents just the conjectural part of a historical reconstruction, when historians image possible events causally or not causally connected with other events in the past and the present. However, future and past are both physically inaccessible and there is no epistemological difference in thinking about making prediction in the past and in the future, if by prediction we mean a description of an event *before* it is known to be happened, that is a novel phenomenon. The only difference is in the control or confirmation of that prediction. Predictions in the future can be considered true if that event they predicted happens. In predictions in the past, observing the predicted event is not possible, but we can find indirect traces of it in the past, present and future. In other words, a prediction in the past has always an indirect confirmation.

There are no deductive predictions of the past, but only probabilistic. This means that any event and period is reconstructed on the basis of a probabilistic prediction. A reconstruction is a prediction in the past that receives sufficient proves in the present to be accepted with a certain degree of probability. While in history all predictions are probabilistic, in science we can have either probabilistic or deductive predictions, like in mathematics.

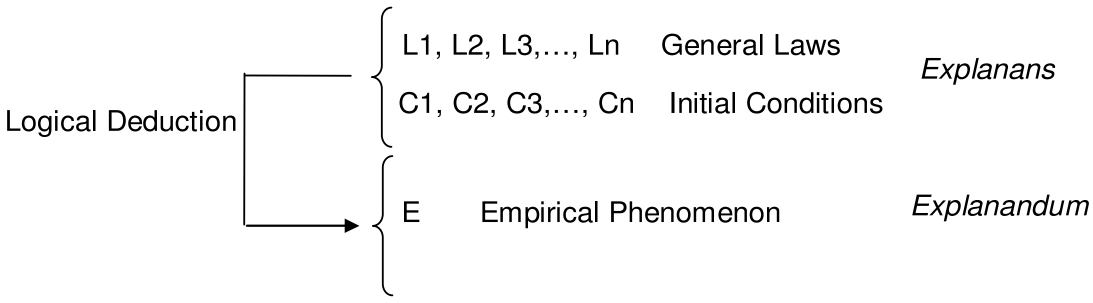
Laplace images an intelligence that could know anything that for us is true only with a certain probability. Note that in his image he makes no distinction between past and future, considered the two world inaccessible to the "certain" knowledge.

> Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it – an intelligence sufficiently vast to submit these data to analysis – it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, and the past, would be present to its eyes (Laplace 1814, p.4)

7.2 The Historical Explanation

While Neurath proposed his unification of sciences based on an inductive epistemology, Popper and Hempel believed in a deductive epistemology and in a model of explanation in which the explanandum is explained by a deduction from general laws and initial conditions. In this model there are general laws L1, L2, L3,…LN, initial conditions C1, C2, C3,…, Ck, and an event of a type E. The explanation of an event of type E is the discovery of causes (in a causal explanation) or conditions (in a non-causal explanation) sufficient to determine E. This discovery can be done only by knowing some general laws, that link deductively or with a certain probability the event with the initial conditions.

This is the explanation model in its simplified form:



This idea of explanation as the established relation between an event, the explanandum, and another event or a set of events, the explanans, comes from Hume and can be found inside the all empiricist tradition. Hume claimed that the explanation of an event can be obtained by means of the observation of that event associated to another event. This regularity, uniformity in the observation of an event with another event constitutes a “generalization”. Therefore for Hume we obtain an explanation of an event when we find the generalization to which it belongs.

A historical explanation of E consists in determining the initial conditions related to the happening of the event E. Initial conditions represent either the causes - in a causal explanation - or other simple events related to E. E is not really the *single event*, since to represent the single event in its infinite aspects would be impossible. E represents the *type* by which the single event can be represented. Both E and C refer to a ‘type’, a set of aspects and properties of events. For Hempel historical laws are necessary to the determination of the initial conditions, since these laws establish that specific events are *usually* correlated with specific initial conditions. Popper, however, denies historical laws and for him a historical explanation should be focused on the research of initial conditions, by using laws from other disciplines, when needed. Hempel and Popper agree on the logical structure of the scientific model of explanation and both see in this model the unification of the scientific method. However, while Hempel considers that the model is one and one is the use and interpretation of it in any discipline, for Popper the unity of the logical structure of the model is compatible with different interpretations of it. For we have the *theoretical* or *generalizing* sciences, the *applied* sciences, and the *historical* sciences. The former are sciences that have the aim of finding and testing theories. Applied or technological sciences have the aim of using the theories to make predictions. Finally, historical sciences aim to explain singular events by establishing the initial conditions.

Hempel claims that in the DN model of explanation general laws connect the explanandum with possible initial conditions, that are the causes of the event in a causal explanation. There is no explanation without a law. Hempel believes that general laws, even if probabilistic, exist in historiography, therefore historical explanation are possible. For Popper, however, according to the idealist philosophy of history, general laws in history do not exist and in history there are only *interpretations*. If the choice of initial conditions are not based on general laws, on the basis of what should a historian make this choice? At this point the answer of Popper is that there is a *trivial general law* which says that a “sane” person as a rule acts more or less rationally. Therefore the logic of historical explanation becomes a *logic of the situation*, the logic of the action of individuals in a specific situation.

> We may say that our actions are to a very large extend explicable in terms of the situation in which they occur. Of course they are never fully explicable in terms of the situation alone; an explanation of the way in which a man, when crossing a street, dodges the cars which move on it may go beyond the situation, and may refer his motives, to an ‘instinct’ of self-preservation, or to his wish to avoid pain, etc. But this ‘psychological’ part of the explanation is very often trivial, as compared with the detailed determination of his action by what we may call the *logic of the situation*. (Popper 1945, p. 97)

Initial conditions describe personal interests, aims, and other situational factors. Historical events are *unique*, in the sense that their explanation depends on an infinite number of different initial conditions, specific for every single situation. The choice of initial conditions is not objectively established by the use of general theories, but it is subjectively based on the point of view of our explanation, on the aspects of a specific event we are interested to. For this reason, Popper prefers the term 'interpretation' for the historical inquiry. It seems that for Popper historical sciences are intrinsically subjective, in the sense that they can not be objectively justified.

Popper is against general laws in history, and in particular he refuses the theory of *psychologism*, the theory which says that there are general psychological laws concerning the rationality of 'human nature'. For him, like for Weber, the terms 'rational' and 'irrational' behaviour have a meaning in accordance with the logic of a specific situation. A psychological analysis of an action in terms of rational or irrational motives presupposes that we have previously developed some standard of what is to be considered as rational in the situation in question.

Even Hempel takes in consideration the peculiarity of the historical event, which can be described in many different ways. However, he does not consider the nature of the historical event as a limit for the objectivity of the historical explanation. A complete description of an event is impossible, therefore the explanandum can describe only one aspect of it. The explanation is objective, but it cannot be *complete*. If the description is not complete, *a fortiori* the historical explanation of it cannot be complete. Any historical explanation is incomplete and also contingent, since the general laws represent points of view that can be changed in the time. Besides this, any generalization that is empirically testable produces an explanation. Only in the case of a generalization that is not empirically testable, we should speak about *interpretation*.

The historian must make a selection of events. He must select what he considers basic for his particular history. This does not necessarily mean that the history will be subjective and arbitrary. If that selection is justified, then history should be considered perfectly objective. This is not an issue specific of historiography, but of all types of knowledge. Knowledge is objective *only for* some determinate contexts and times.

In conclusion, for Popper what we can reach are only many historical interpretations. For this reason, he suggests for history an epistemological *pluralism*. We should note that Popper view is not quite close neither to Neurath's view nor to Hempel's view. Instead, his conception of history is closer to the idealist philosophy, in the sense that he reaffirms a dichotomy between disciplines, even if not between natural and human sciences. He separates *historical* from *generalizing* sciences and to justify this separation he uses an argument based on the nature of the historical evidence. The information about events we want to explain in history is often very poor and facts are selected in accordance with subjective points of view. This selection would not be a problem if we could finally reach justifiable theories, but it is a problem since facts are selected to fit in the theories and theories can not be independently justified, but only in a circular way, by using the same facts that have been initially selected to be explained. Therefore in historical sciences only interpretations of single facts are possible.

> For in history (including the historical natural sciences such as historical geology) the facts at our disposal are often severely limited and cannot be repeated or implemented at our will. And they have been collected in accordance with a preconceived point of view; the so-called 'sources' of history only record such facts as appeared sufficiently interesting to record, so that the sources will, as a rule, contain only facts that fit in with a

> preconceived theory. And since no further facts are available, it will not, as a rule, be possible to test that or any other subsequent theory. Such historical theories can then rightly be charged with being circular in the sense in which this charge has been unjustly brought against scientific theories. I shall call such historical theories ,in contradiction to scientific theories, 'general interpretations'. (*Ivi*, pp.265-266)

Finally, a historical reconstruction cannot be falsified, because it is only an interpretation of single facts. Since historical sciences do not admit falsification, their predictions are nothing more than *prophecies*. Another problem for Popper, which is not specific of historical sciences, it is the problem of probabilistic explanations and predictions, the issue concerning probabilistic hypotheses and their falsification. Falsification of historical hypotheses is difficult also if "reinterpreted" in a pragmatic way for probabilistic hypotheses. First, this probabilistic version of falsification makes sense in a Bayesian framework, that is when we consider the posterior probability of a hypothesis. Also in this framework this idea had in the past several problems[9], since events that could falsify a historical reconstruction must be either reproducible - if they have already happened - or they must have a *not very low* probability to happen, sufficiently low to be considered *impossible*, if they did not happen yet. While it is not possible to reproduce a historical event, it is difficult to define objectively when an event can be considered possible or impossible, when we work with many different reconstructions, which can have in most cases a very low probability to be happened. Falsification does not make sense in a likelihood theory of evidence, since likelihood represents the support that a hypothesis gives to the evidence, and not the reliability of a hypothesis given the evidence. Finally, likelihood has a comparative meaning which does not fit the idea of falsification of a single hypothesis. When we use likelihood to

---

[9] See about this topic Howson and Urbach 2006.

valuate a hypothesis H, it makes no sense to compare H with *non*-H. Sober (forthcoming) suggests an updated but weaker version of *modus tollens* for probabilistic hypotheses, where the conclusion "*non*-H" changes in "*not*-O is an evidence against H".

$\Pr(O \mid H)$ is very high.

$notO$

================

$notO$ is evidence against H.

In conclusion, for Popper, even if he claims that the model of explanation is unique for all sciences, there is more than a difference in the interest between historical and generalizing sciences. In history general explanations are not possible at all, but only interpretations of single facts, always subjective and charged of points of views used in the selection of facts. Moreover, there is not in history the symmetry between explanation and prediction, since predictions are based on the possibility of general laws. We have seen that for Neurath in historiography and social sciences the explanation was possible but the prediction presented some difficulties. For Popper in historiography no general explanations and predictions are possible, but they are in social sciences.

> […] there can be no history of 'the past as it actually did happen'; there can only be historical interpretations, and none of them final; and every generation has a right to frame its own. But only has a right to frame its own interpretations, it also has a kind of obligation to do so; for there is indeed a pressing need to be answered. We want to know how our troubles are related to the past, and we want to see the line along which we may progress towards the solution of what we feel, and what we choose, to be our main tasks. (*Ivi*, p.268)

Popper seems to hold a *pragmatic* position about the historical sciences. While historiography can not produce general theories, historical interpretations can still help the progress of the society, whether we avoid a-critical behaviours in selecting facts.

7.3 Dray and the Model of "Rational Explanation"

The DN model of explanation has been very discussed in philosophy of history. The main problem for many authors was the problem of accepting general laws in history. We have seen that while in the idealist tradition the issue is the different object, which represents a limit to generalization, in Popper the issue is the different aim, since history aims to explain particular facts and the poverty of the available sources. While there is a position of total nihilism towards generalizations in history, there are also intermediate positions. These positions consider generalizations to be necessary for the understanding of historical structure, but they claim that any historical generalization must be considered contingent. Joynt and Rescher (1961), for example, propose the idea of *particular generalizations* or *contingent regularities*, where general laws in history are designed for the interpretation of a particular period and they are expressed according to a specific technological level, cultural context, etc..

Inside the historical nihilism, Dray (1967) rejects the DN model of explanation and proposes the so called "rational model". His argument is that history is the theatre of human actions generated by reasons. The DN model is designed to be applied to causes and not to reasons, therefore it cannot be used for historical explanations. The argument is based on the difference of objects, that is, exactly like in Collingwood, the world of thoughts. However, in Dray there is a new element. History is not only the world of thoughts, but the world of "rational thoughts". The model of rational explanation is founded on the assumption that "all men are rational". In other words, Dray does not eliminate all the generalizations from the historical method. Any action must be explained through the calculus that the agent made in order to obtain a certain result, in a specific situation. The idealist theory of empathy becomes in Dray the theory of the *rational calculus*. The explanation based on this calculus is called rational explanation and aims to show the rational foundation of what people did in the past. The rational foundation is called by Dray "what must be done" in a specific situation, for that aim and those reasons. The "to be done" is for Dray objective, given a specific situational context and given the assumption that the agent in that situation acted rationally. The "to be done" is for Dray the result of a *principle of action*, which is a *principle of rationality*, and it does not represent a generalization. But some generalizations are hidden in this principle, and they concern the fact that all people are rational and rational people act rationally in any situation. But, why should we assume that all men are rational or that, given they are rational, they act in a rational way? And in particular, why should we assume that that specific man acted in a "rational" way in that situation and that all men in that situation would act in the same "rational" way? How must be established what is the "rational" behaviour for a specific situation? Dray's rational explanation is not objective like it seems to be and it implies some kind of *evaluative criterion* or rationality. The assumption of rationality is the same we have seen in Popper's logic of the situation, where a trivial law is accepted as true: all sane people act usually more or less in a rational way. This criterion is subjective and *a posteriori*.

Hempel replied to this model by claiming that a rational explanation is not even an explanation, since to show that an action is appropriate in a certain situation represents only a justification of that action, but it does not explain *why* it has been made. In a rational explanation, explanation and justification of the action are assumed to be the same thing. However, a rational explanation is an explanation *only if* one assumes that the agent has a disposition to act rationally in a given situation.

7.4 Explanation and Causation

Often in history an explanation is considered to be a causal explanation. We have seen that in historiography there is a strict dichotomy between a casual explanation of events and an interpretation or explanation of actions trough the reasons behind them. In Collingwood, Gardiner, Mandelbaum, Dray, Popper, for example, history is based on the interpretation of the initial conditions, which are the purposes, intentions, and reasons behind the actions of individuals.

In Weber "explanation of an event" means to find the causes that can be considered *adequate* to produce that event. He focuses on every *single* event that can be taken as *sufficient* for the happening of another specific event. Weber applies a *by elimination* model in historiography. He images ideal worlds in which history could have been different, by eliminating the event that has been determined as one of the sufficient causes of some other events. He tries then to analyse the possibility of alternative representations of the same history.

In the DN model of causal explanations, general laws make the initial conditions to be causes of the explanandum, instead of sufficient conditions to the explanandum to happen. In other words, the hypothesis that in a certain context C1, C2, C3,…, Cn cause E represents the explanation of E. From the fact that in history a complete explanation is not possible, it follows that what we obtain many *sufficient* explanations which aim to explain certain aspects of the events. There are an infinite number of initial conditions we can consider as causes for the infinite number of aspects of the explanandum. Marc-Wogau (1962) and Scriven (1966) establish a *continuum* respectively between necessary and sufficient reasons and cause and effect. For Marc-Wogau the *continuum* ends up with a solution *ex post factum*, in the sense that a reason is only sufficient before the agent acts and it becomes necessary after the action. The single reasons are the causes of the single fact and they are necessary after that the action has been done. In my view of history, what he calls necessary in a metaphysical sense, it is only a sufficient reason in our reconstructions of history. Initial conditions are all sufficient and they are part of our representations. However, the concept of "necessary" is often used in historical reconstructions, and it reflects a deterministic interpretation of the nature of historical processes. Consider a simple process in molecular evolution: a mutation arises by chance and it is fixed in the genome. In the famous Monod (1970) idea of evolutionary process, we find the paradigm of the "chance and necessity". The idea is that a mutation, after being randomly generated, it can be selected and from that point it becomes part of a necessary historical pattern, which shapes in a deterministic way the whole evolutionary process. However, the evolutionary process is not necessary in a historical sense: nothing that happens in nature should be considered like something *could not be otherwise*. In a specific ecological context, more than one possible pattern can be followed by nature. When a specific variation has been selected and pushed under functional constraints, we claim that it is part of a necessary functional pattern. There is a time *t* after which we decide that the random variation has become part of a deterministic process. Obviously, we do not know exactly when this mutation reaches the time *t* and by convention we decide that it is when it becomes part of a network of relations "complex enough" that any part of it is needed for the function of the whole. But it is a functional necessity, established in our network-representation of the biological process and it is not a historical necessity!

While Marc-Wogau works on the distinction between sufficient and necessary reasons, Scriven argues that historians choose their own language, which can be or not a causal language and also, in the case they choose a causal explanation of events, there is a *continuum* between cause and effect, since only the *context* the historian chooses in the explanation can transform an event in a cause or an effect. This idea recalls the pragmatic model of causal explanation considered in philosophy of science by Van Fraassen (1980). According to Van Fraassen, a causal explanation is an answer to a question Q of the form: Why P? where P states the fact to be explained, i.e. the explanandum. He calls these questions "why-questions" and argues that they are characterized by three elements: the theory, the fact, and the context. What the cause and what the effect are in an explanation changes in function of the context. Moreover, the why-questions are necessarily contrastive, for example: why this rather than that? The causal question must be *contrastive* to be unambiguously answered; this means that a non-contrastive question can have many different relevant answers. Therefore, questions like "Why Caesar crossed the Rubicone?" have many answers, depending on the hidden alternative we are considering, like "rather than staying in Rome" or "rather than crossing another river", etc..

In the historical reconstruction, the causal explanation of events should not be taken in a metaphysical sense, but considered as a possibility in our representing history and strictly contingent, in the sense of being context-dependent. Therefore we have a causal explanation when the question: Why has the event-explanandum happened? Is interpreted as: What are the causes of the event-explanandum? We can imagine that in our reconstruction one or more events can become sufficient causes in the specific context of other events we are considering. These other events, which could be named "secondary" causes, constitute the context that makes that event to be the sufficient cause in our reconstruction. This reasoning can be the same if we use the term ‘reason’ instead of ‘cause’. In this case we consider a possible reason for which the agent acted in a certain way. This reason, without the context of desires, solicitations etc. that we are considering in our representation of that historical moment, is not even sufficient. In this case we have that explanation which we have called above “rational explanation”. Many philosophers have studied the possibility of reconstructing history by reconstructing the reasons of the agents, like for example D. Davidson, who speaks about *pro-attitudes* to certain kinds of actions, which can become reasons in specific situations[10]. This shows that historical narration can also be an explanation without necessarily be a causal explanation. Like we have seen in Dray’s philosophy of history, the “calculus of the agent” represents the calculus of the agent’s reasons and intentions, which historians should evaluate in order to explain his/her action. “Why did he/she do that?” This question, that for many historians represent the typical historical question, has an answer in the context of reasons of the agent. I would like to stress again the position we have seen in Carr, Croce and Collingwood, for whom the context of any interpretation of an event is not only the context of the agent acting in the past, but it is also the present social, cultural context in which historians investigate the past. Therefore, a reconstruction is a representation made in a context of sufficient causes or reasons, which also represents the answers to the questions that historians formulate in their social, cultural context, on the basis of their interests.

In conclusion, our reconstructions of the past should be considered as historical representations with no metaphysical implications. In particular, I shall argue that representations of phylogenies are *images* of the biological past, based on probabilistic models and justifiable only inside what one could call a “mode” *á la* Oakeshott, in the

---

[10] See about this topic Davidson 1963.

sense of a coherent world of concepts with its own ontology and epistemology. The ontology is made by the probabilities of the biological entities, and the epistemology is made by probabilistic theories of confirmation and prediction. There are no naïve reconstructions, but only representations that have a specific significance inside different probabilistic models.

7.5 Structures in History

How do we represent history? We can follow the idea of M. Stanford, who in his book of 1986 on the historical knowledge denies like Popper the existence of laws and theories in historiography and suggests to use, rather, the concept of structure. Roughly, he defines a 'structure' as a set of relationships between elements which can determine the character of the whole. Stanford argues that in epistemology of history there are two different structures: the inner structures and the imposed structures. While the first ones are found in the world, independently of human will or choice, the second are imposed by the historians in order to interpret and understand the world, therefore they have a human origin. The imposed structures are made by our languages, social behaviours, beliefs. Structures derived from historian's own social experience seem to be imposed for two reasons: understanding history and enjoying the fiction, which is part of our work of reconstructing history. The historical hypotheses are to be formulated inside conceptual structures and social structures, and the whole resulting structure of history could be considered similar to narrative poems or images, which will constitute new historical material for the historians of the future. The structures inherent in the world are represented by the historical evidence. The activity of knowing history involves the historical evidence; the construction in the historian's mind; the influence of the social mind, that is the society in which the historian works; the previous interpretations of history; all this structure produces new historical actions, that is new history. The scheme below shows a possible structure of the historical knowledge.

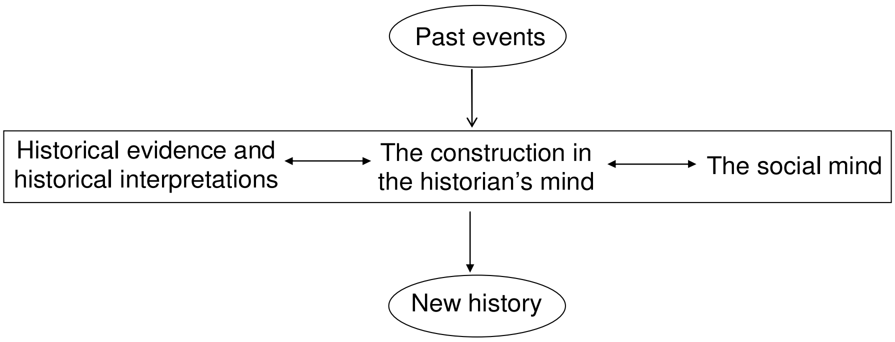


Figure 19. The Structure of the Historical Knowledge. The Arrows Indicate Relationships of Cause and Effect among Different Structures.

The social and conceptual structures and the structure of the historical evidence - that is the evidence read in a historical sense - produce different images of the past events, different images of history. At the same time they represent a historical action, which is going to represent the "new" history to be interpreted by other historians in the future.

CHAPTER 3

IMAGES OF THE PAST

1. Understanding the Biological Past

How do we understand the biological past? If we believe that the theory of evolution is true, first of all we should consider that our understanding of the biological past will be inside the evolutionary framework. Any event we consider in the biological past is understood inside this framework. It is trivial to say that we cannot observe events in the past, but the evidence – effects, traces, remains – they are left in the present. Therefore, understanding the past is always a reasoning from the present evidence to facts or events in the past. It concerns predictions on the past on the basis of the present evidence. Hypotheses about the past involve correlations among observations, inside different conceptual frameworks about evolutionary processes. In a definition of "historical sciences", Tucker (2004) distinguishes them from other sciences by the evidence they focus on, which always includes correlations or similarities between documents, testimonies, languages, or species. These sciences attempt to infer the causal origins of these correlations or similarities. But evidence does not say anything to us, if we do not make some assumptions about processes that made this evidence possible. Without assuming some history to be true we can not predict the past from the present. Our understanding of the past is probabilistic, therefore our predictions on the past are probabilistic predictions. In phylogenetic inference probabilistic theories are used to associate to different possible pasts their probabilities to be true. It seems odd to speak about the past in terms of prediction and patterns coming true with a certain probability, but as I have said in chapter 2 (§7.1), it depends of our definition of the term ‘prediction’.

When we aim to understand similarities inside the theory of evolution we have necessarily to consider evolutionary processes that created these similarities. Phylogenetic hypotheses without assumptions on evolutionary processes are undetermined. Sober (1988, 1990) pointed out that one can obtain different probabilities (likelihoods) for different histories only under the assumption of different evolutionary processes. This is the famous philosophical problem stated by Duhem’s thesis, according to which an experimental test always involves several hypotheses. What we have access to is the probability of the conjunction (H & A), where H represents the phylogenetic hypothesis and A represents the assumptions about the evolutionary processes. Unfortunately, these processes are unknown. Biologists use different stochastic and not stochastic models based on several assumptions in order to make predictions from the present evidence to the past. But, are these assumptions realistic? Or, better, consistent with our current knowledge about evolutionary processes, what we actually know about them? For example, we know that evolution is not made of only neutral mechanisms, but most of our models give the same weight to different kinds of mutations, different characters and different sites. We know that evolution does not follow any criterion of “optimality”, in the sense that no optimal variations in different ways characterize the direction in which evolutionary changes occur. However, we do use models based on criteria of optimality like Maximum Likelihood and Maximum Parsimony methods[1].

In order to understand better the problem of the conjunction “history and its model”, consider this simple example. There is a person with no memory of her past.

---

[1] These two methods are called optimality methods since they permit to choose the tree that optimizes particular values, like the value of minimum number of change and the value of likelihood.

She can make several conjectures about the patterns that have lead her to the life as it is at the present time, and all of them have the same probability to be true, in the case she has no information about the particular decisions she has made in the past. In this special case to decide for a pattern that explains the evidence by using probability would be impossible, since any historical hypothesis would turn out to have the same probability, in the sense that the evidence would be totally independent of the past and free to be explained with any kind of history. In probabilistic terms, any profile would give the same probability to the evidence. Ironically, this is exactly the case of phylogenetic inference. The process that links the past to the present is unknown and phylogenetic hypotheses need some assumptions on it.

2. The General Structure of Phylogenetic Inference

The structure of phylogenetic inference involves a hypothesis of common ancestry with assumptions on the ancestor; models characterized by assumptions on the evolutionary process; a choice of the evidence considered relevant to make predictions to the past. In this chapter I am going to show that a historical reconstruction is better seen as an ideal representation or "image" of the past, instead of as a reconstruction of what happened. Process models constitute "ideal models", which make possible our images of history. They are "ideal" since they are made by a set of ideal assumptions on the evolutionary process, that is the whole historical process that links the past to the present.

Evidence + background assumptions + ideal models → Images of History

It is bizarre to consider that without an ideal model of *history*, to make an inference on *history* is not possible. This sentence makes sense because by 'history' in the two cases I mean something different. The first term refers to the history represented by the process that has generated the evidence and the other term refers to the representation of a topology – a tree – which gives us the idea of the genealogical connections between the present and the past. The first history is unknown and must be assumed in order to formulate an inference about the second one.

In order to make an inference about the past we must decide what evolutionary evidence should be taken in consideration and what questions should be asked on this evidence in order to make predictions on the past. However, we shall see that there are several problems regarding what questions we should ask and which evidence should be taken in account. In any method of phylogenetic inference, the choice of evidence involves many considerations and presuppositions about the process. As we have seen in chapter 2 in Weber's methodology of history, in the historical reconstruction the choice of what evidence should be considered epistemologically meaningful depends on the presuppositions that lead our research. This is obviously a very well known idea in philosophy of science, belonging to the whole tradition that assumes a conceptual component in evidence. However, in phylogenetic inference not only the evidence is conceptually laden, but the whole process of historical reconstruction is meaningless outside of some ideal models of assumptions. Therefore biological history, exactly like human history, should not be considered as reflecting the past as it is, but as a representation that gives a historical meaning to the evidence. There are different kinds of assumptions, some of them characterized the general patterns of evolution and they can be considered as "background" assumptions, because of their very general character. There are assumptions on the divergent nature of evolution, the uniform character of change and on a common ancestor. All these presuppositions are part of our historical interpretations of the evidence. The last one is particularly interesting, since it seems to be an essential hypothesis in any historical reasoning. For we shall see that the hypothesis of common ancestry is used as well in methods of reconstruction of histories of languages and texts. In the following sections I shall start with a simple example in philological reconstruction and then I shall move towards phylogenetic inference. The hypothesis of common ancestry is connected by a particular principle of parsimony, the "cladistic" parsimony. This principle is not the same as the principle of parsimony adopted by Maximum Parsimony methods, as I shall show in §2.7. The cladistic principle of parsimony has been a problem in phylogenetic systematics. The problem is whether this principle has metaphysical implications on the nature of evolution. The solution to this problem is that the tree of common ancestry, that is the tree of maximum cladistic parsimony, has a greater likelihood than the tree of separate ancestry. In §2.6 I shall present the probabilistic justification of the principle of common cause, that is the connection between likelihood and parsimony. In the other sections of this chapter I shall give my view of the "historical thinking", that is of genealogical hypotheses. Finally, I shall face the epistemological issue of the requirement of total evidence in phylogenetic inference and I shall suggest a possible solution to the problem.

3. The Principle of Common Origin

3.1 History and the Common Ancestor Explanation: The Case of Philological Reconstruction

Philological reconstruction is a special case of historical reconstruction. I want to present this example in order to show that there is an historical way of knowledge, instead of separate disciplines like history, phylogeny, philology and that this historical way of thinking is based on the hypothesis of common ancestry. In the case of philology, for example, two independent causes for the same evidence imply randomness and if we must assume randomness we would not have any history to reconstruct. Therefore, history must have a “sense” in order to be reconstructed and this sense is inside the hypothesis of common cause. Philology is the reconstruction of the history of a manuscript, from its *archetype* to the present, where the archetype is the first copy of the manuscript. The method of philological reconstruction based on the idea of common ancestry was developed by K. Lachmann in 1800 (see for an analysis of his method Timpanaro 1963). This method has a strong analogy with the method of comparative historical linguistics, used with the aim of discovering a language-ancestor of linguistic families. In both cases the aim is the distinction of elements inherited by innovations and make conjectures on the common origin of manuscripts or languages. The same innovation between some copies of a same text or between some languages of the same family, shows their belonging to the same sub-group: the corruption of a text, it is an innovation in respect of a text previously transmitted, exactly like in the linguistic innovation. Besides all these analogies, there are also some differences. One of this is the fact that a corruption in the history of a manuscript is considered an error, and not only a neutral innovation. In linguistics, when a corruption becomes accepted and part of the language, it is not *ipso facto* considered an error anymore.

Reconstructing the history of manuscripts is a reasoning involving similarities and differences between two or more manuscripts. Consider the case we have two manuscripts A and B and we want to give a historical explanation to the similarities and differences between them. We can either think that the manuscript has been produced in two copies by two different people or that it has been produced once and than copied by someone who made some modifications in respect of the original text. The two historical hypotheses will be:

a. There is only one archetype X modified in two A and B manuscripts

b. There are two manuscripts X e Y generated independently e then modified in two A e B manuscripts.

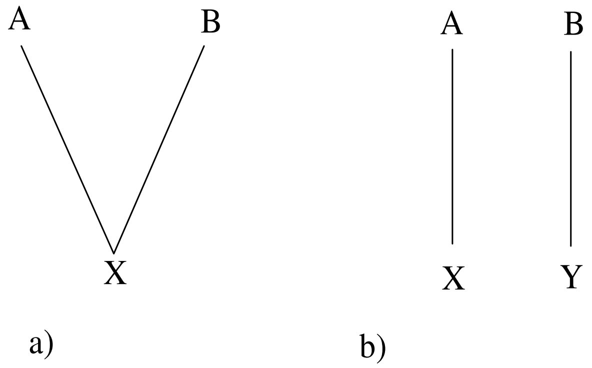


Figure 20. Separate and Common Origin in Philology

The explanation with only one copy is the most plausible, given the most historical processes we could imagine, but not always. Let us assume, for example, extreme cases in which people who make copies of different books share information with each other and make the texts very similar; or a case of a regime in which people who write books have to follow a common standard of stories and styles. In these cases texts would be very similar but not for a common cause. We can just say that the hypothesis a) is more plausible of the hypothesis b) in the most likely historical processes we could imagine. If one thinks about DNA sequences used in molecular evolution, for example, it is not trivial to decide what is the product of evolutionary constraints like natural selection and what can be considered as a modification of an ancestral sequence. In general it is not trivial to understand what kind of information a difference or a similarity of a character can carry on if one does not know anything about the evolution of a character in different organisms.

Therefore the history of manuscripts is based on three general ideas, which are characteristic of the reconstruction of biological history as well. First, that we need differences and similarities in order to reconstruct a history. Second, that we have to assume a process of modification in order to explain the differences and similarities. Third, the common origin is more plausible than the separate one.

3.2 The Choice of Evidence in Phylogenetic Inference

We have seen in chapter 1 that whether we want to use the principle of cladistic parsimony we must consider only apomorphic traits, since – cladists claim – only these traits have a genealogical meaning. What does it mean exactly to choose only apomorphic traits? It means that we have already made an ontological assumption about the process that has generated the evidence. The assumption is that there exists a common ancestor from which traits have been inherited. However, we should note that a process that hypothesizes separate ancestors is genealogical as well and therefore the choice of a hypothesis of common ancestry can not be justified by saying that it represents the only genealogical hypothesis. The hypothesis of common origin can be justified only by saying that a common cause explanation of the evidence should be considered more plausible than a separate causes explanation. Since phylogenetic theories are probabilistic, by ‘plausible’ I mean more probable (posterior probability) or likely. The hypothesis of common ancestry is the most plausible and therefore it is justified only in certain evolutionary contexts, which can be assumed but not be tested empirically. Therefore this hypothesis will be valid only inside certain evolutionary contexts, where it can have a probabilistic justification. The context involves two things:

the *process of changing* that occurred in any branch of a tree, and the *character state of the ancestor*.

When we aim to reconstruct the phylogenetic relationships that connect different species to each other, generally we prefer hypotheses that allow us to interpret similarities as homologies. For example, when we consider a group of species with and without a particular character like wings, we can choose between two different ways of explaining these similarities. One can explain them by saying that the shared characteristic is a homology inherited from a common ancestor, or by saying that it is a homoplasy that it is evolved independently in the separate species. Common sense and scientific practice both embody a preference for the common cause explanation. This preference has a probabilistic justification, if the processes occurring in lineages obey a principle that Sober (1988, 2002c) called the "backwards inequality" and says that:

(1) The likelihood of the hypothesis of the ancestor with wings is greater than the likelihood of the hypothesis of the ancestor having no wings on the same evidence that descendants have wings.

He argues that the backwards inequality is independent of the question of whether stasis is more probable than change, that is instead represented by the following inequality:

(2) The probability of the descendants having wings given the hypothesis that the ancestor having wings is greater than the probability of the descendants lacking wings given the ancestor having wings.

This inequality is considered to be true only for some traits and in some lineages, therefore it is not a very robust statement. The "backwards", however, is believed to be much more robust. Some questions should be asked: Why should (1) be true and much more robust than (2)? On the basis of what evidence? Yet, even if (1) is independent of (2), does it follow that (1) has no implications on the process of changing? Let us see what a backwards inequality means and whether it is really "neutral" towards the evolutionary patterns of changing, in the sense that it can be justified in any evolutionary context. In the backwards inequality we compare two hypotheses about the character state of the ancestor over the evidence about the character state of the descendant and we claim that the conjunction (1 & 1) – where the first number refers to the character state of the descendant and the second to the character state of the ancestor – is more plausible than the conjunction (1 & 0), given different sets of characters under consideration. When we express the inequality (1) in this way it has exactly the same meaning of the inequality (2), that is that a history with no change supports the evidence in the present time better than a history with one change. Is this always true? Not really, since it depends on the evolution of the characters we are considering in our analysis. The "backwards" inequality is not independent of the question about stasis and change. This illusion is maybe given by the fact that in (1) we are considering a hypothesis about the state of ancestral characters instead of a hypothesis about the character state of the descendant and what belongs to the past can not be influenced by a process which will transform it in something that belongs to the present time. This is because this process did not happen yet. However, this way of seeing the issue is misleading, since our historical representations are in the present, they are pictures, interpretations of the present evidence, which is given by the characters of the descendants, with the support of ideal models in which we have both the hypothesis concerning the character of the ancestor and the process of changing. In both the disjunctions we have a different conjunction between past and present traits and the conjunction implies a reasoning about the evolution of those traits in both cases, since the conjunction is the product of this evolution.

3.3 The Principle of Parsimony in the Hypothesis of Common Ancestry

The principle of parsimony is used in *two* different ways in phylogenetic inference; we can say that there are two concepts of parsimony. The first one is related to the idea that history is constructed on the common ancestry hypothesis and it is exactly the way in which it is used in the method of cladistic parsimony. For this hypothesis in the cladistic school is *needed* to reconstruct monophyletic groups and therefore it is *needed* to make a genealogical reconstruction. This use of the principle of parsimony leads to choose the tree characterized by the minimum number of homoplasy. In a probabilistic framework, cladistic parsimony is not the method that makes impossible some events, but the method that prefers the phylogenetic hypotheses that minimize the number of improbable events. This use of the principle of parsimony has been justified by means of the likelihood (see below).

However, parsimony can be used in another way, inside Maximum Parsimony methods, in which the aim is not only to minimize the number of homoplasies and therefore to choose the hypothesis of common ancestry over the hypothesis of independent ancestry, but the aim is to choose between several hypotheses the one that can explain the same evidence by postulating the minimum number of changes. This method does not consider the probability of changes, but makes the fewest the number of changes in a tree. Figure 21 shows the difference in the two concepts of parsimony. There are two phylogenetic hypotheses A and B, representing the same topology but with different evolutionary events. The first hypothesis indicates 3 changes in total, happening in the middle of the tree, corresponding to three multiple originations. The originations on the right give finally rise to 16 homoplasious matchings between the tip species. The hypothesis B indicates 5 changes but only 6 homoplasious matchings. Therefore, hypothesis A is more parsimonious than B following the second concept of parsimony, but it is less parsimonious in a cladistic sense.

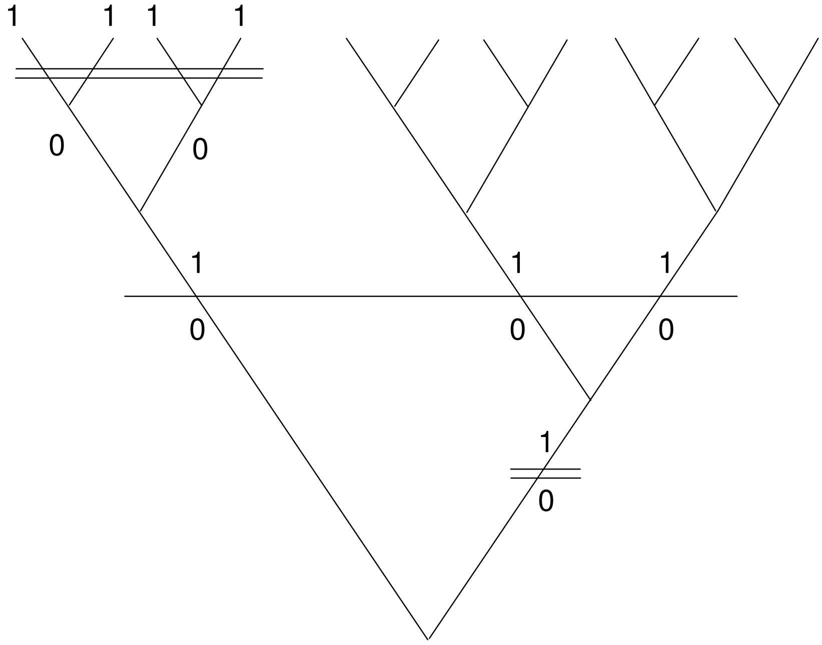


Figure 21: An example in which the two concepts of parsimony are different. Hypothesis *A* postulates that character states change from 0 to 1 at an intermediate time indicated by the single line. Hypothesis *B* hypothesizes that the changes occur much later on the left branch, and earlier on the right branch, as indicated by the double lines. Hypothesis *B* postulates more changes, so *A* is more parsimonious than *B* in that sense. But *B* has greater *cladistic* parsimony because it postulates fewer homoplasies. *B* loses 4 homoplasies on the left branch, but gains 16 on the right branch.

The second use of the principle also applies in the explanation that we give to the observation of a character of a species in state 1, given the assumption that the character state of the species-ancestor is in state 0; in this case we say that *only one change* has happened long the branch and not multiple changes. This very general use of the principle of parsimony is expressed by saying that we should postulate that nothing happened if we cannot demonstrate the contrary. Look at Figure 22, which explains simply what it means to choose the shortest pattern in order to explain the same genealogical relationship between two species. We can postulate an infinite number of changes and intermediate ancestral species, but usually we postulate one change and no intermediate species.

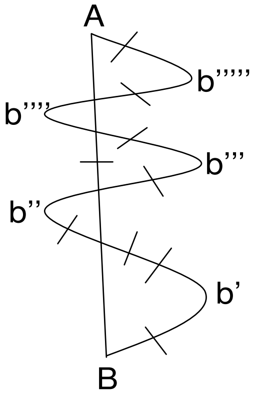


Figure 22. The Pattern of Maximum Parsimony

Note that when we speak about homoplasies and homologies, we are referring to our interpretations of particular *matchings* of characters. Whether a character is homoplasious depends on the objects whose matching is at issue and not on the characters itself. It is important to stress this point, since we shall see in §7 that the evidence in phylogenetic inference is given by matchings between characters and this is going to be a strong epistemological limit for this historical discipline. It is very plausible that if we started to look at the correlation in the evolution of the character states, the meaning of what should be considered homoplasy and homology would change.

Any principle can be used in a global or local sense. We can either decide that a principle makes sense in *any* context or that it counts only in *some* specific context. This is true also for parsimony.

Most attempts to explain the epistemic relevance of parsimony treat the problem globally. They assume that if parsimony is epistemically relevant across a range of inference problems, that it must have that relevance always for the same reason. However, it is worth pondering the possibility that the justification for using a principle of parsimony may vary from problem to problem. Perhaps parsimony needs to be understood locally, not globally (Sober, forthcoming).

What is a "global" principle? Understanding and using a principle as a global constraint on inquiry implies the idea that it is *a priori* justified, in the sense of logically necessary. Otherwise, whether we consider that principle as a local constraint, we must be able to find a justification for it, in the sense of finding a reason why in a particular context it is connected with the plausibility of a hypothesis. The use of parsimony in phylogenetic inference can be considered local in the sense that it is justified only in certain contexts. But it is controversial if this principle, that is applied in the homonymous method of phylogenetic inference, is based on presuppositions about the nature of evolution and in this case if these presuppositions are plausible and fully justified[2].

3.4 *A Priori* Parsimony: the Principle of Simplicity

In epistemology, the principle of parsimony can be expressed as valid a priori and has metaphysical implications. Such a principle is very old, as much old as the principle of causality. Today is widely known as the "Ockham's Razor". Ockham had expressed it in different ways. One of these ways is: *Pluralitas non est ponenda sine necessitate.*

---

[2] Here we are implicitly assuming that the principle of simplicity is chosen in science because it is epistemologically relevant and we are not considering other possible reasons (i.e., esthetical).

An interpretation of it with metaphysical implications is: “catch the essential in the real and construct theories capable of rendering every experience redundant” (Sober 1975, p.1). A non-metaphysical interpretation is: choose the simplest theories that can explain the reality. The principle of simplicity does not need to be taken metaphysically, but only as a methodological principle of organizing reality in our theories and as an epistemological principle in the sense that it can improve our knowledge of the reality. To say that we eliminate whatever is redundant in the world in order to have a theory reflecting the essence of the real represents a naïve realism. In this sense we justify the use of the principle only if we new that in the nature there is a *simple core* that we can catch by eliminating the *noise facts*.

In alternative views of this naïve realism the principle of simplicity can be interpreted in the sense either of an “illuminated” realism or in an instrumentalist way. In the former position we use simple theories in order to know the reality, in the latter the knowledge of the reality is not needed. In both positions we are interested in how useful a simple theory can be instead of a more complex theory in improving our “control” of the reality, like predictive control. In this sense we could be interested in how much more accurate a simple model is in making predictions compared with a more complex model. This is what we have seen about the Akaike framework, in which simplicity and fitting of data are weighted in their contributing of *predictive accuracy*. In this criterion it is expressed by the number of adjustable parameters of a model and it is considered as a property of a model that increases its predictive accuracy; therefore, it is justified from an epistemological point of view.

The choice of the principle of simplicity in theories must be an epistemological choice without metaphysical implications. Its justification is not in the simplicity of the reality, but in showing that in some ways simplicity can contribute to the plausibility of the theories, for example by means of predictive accuracy of simple theories or by means of the fact that in some contexts simple theories are more likely to be *true* than complex ones. Sober (1975) suggested this use of probability in order to justify simplicity. Inside the “mode” of probability, where the term ‘mode’ is used in the Oakeshott’s sense of a conceptual world characterised by a specific epistemology and ontology, simplicity has not a metaphysical sense anymore but it is an epistemological concept; in this mode of experience simple theories are justified over the more complex only if they are more likely to be true than complex ones. In conclusion, the local meaning of parsimony has an epistemic value and therefore it must be related in some ways to the plausibility of a hypothesis.

3.5 Local Parsimony: the Case of Cladistic Parsimony

How belong the principle of parsimony in cladistic parsimony criterion?

Any hypothesis of homology implies parsimony. For, to explain similarities among species by postulating a common ancestor from which they have inherited those traits (homologies) is more parsimonious than by postulating two independent originations (homoplasies).

In the phylogenetic inference, to hold this principle means to prefer the genealogical reconstructions that minimizes the number of homoplasies (evolutionary parallelisms and convergences) required by a tree.

Consider the abstract example illustrated in Figure 23, in which there are two taxa that share the same character X.

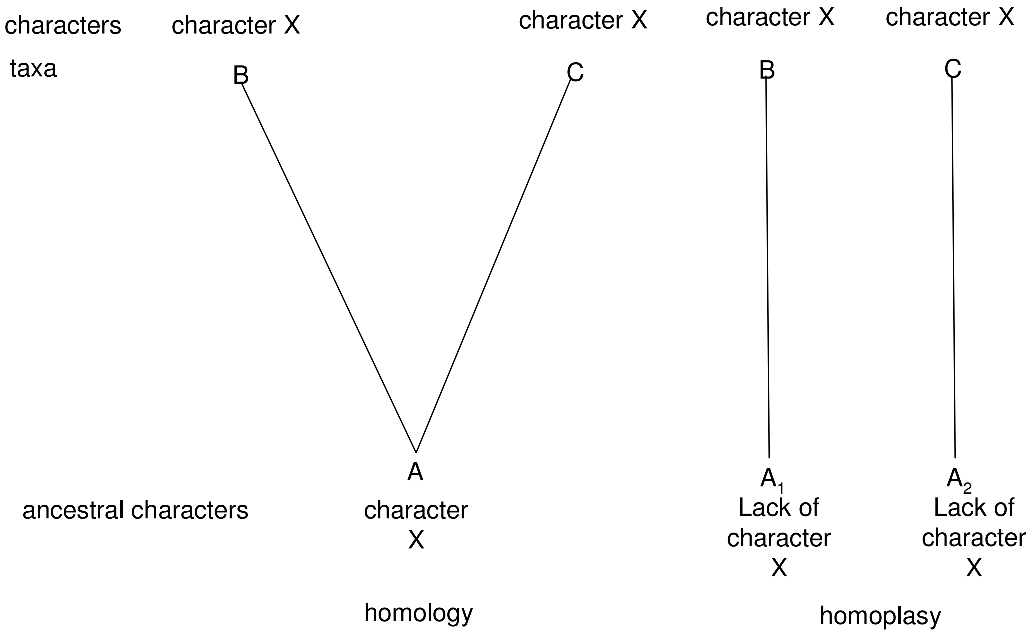


Figure 23. The Principle of Cladistic Parsimony

How can we explain this similarity? In two different ways. We can postulate that they have inherited the character X from a common ancestor. In this case we would consider this similarity as a homology. Otherwise, we can postulate that they have evolved this character independently in two different lineages. In this case we would consider this similarity as a homoplasy. The first hypothesis appears to be the more parsimonious because it postulates one origination event. Homology can be seen also as the hypothesis of common cause explanation and the homoplasy as the hypothesis of separate cause explanation.

How might we decide among this two kind of hypotheses?

Cladists have accepted the first one because it is the most parsimonious. Some of them (Gaffney, 1979; Eldredge and Cracraft, 1980; Nelson and Platnick 1981; Felsenstein 1973, 1978; Farris 1973) have tried to justify parsimony by using Popper's idea of falsifiability. They argue that a phylogenetic hypothesis is falsified each time it is forced to say that a character evolved more than once. They conclude that the least falsified hypothesis is the one that requires the fewest number of homoplasies. Other cladists as Farris (1983) tried to justify the use of parsimony to infer genealogical patterns, arguing that parsimonious trees do not require any assumption about the real evolutionary process.

Another way used to decide between the two different hypotheses - and in particular to justify the use of the most parsimonious tree - is the Likelihood Principle. The argument – suggested by Felsenstein in 1973 - is the following: to use parsimony does not require that the world is parsimonious, since the rarity of change *suffices* for parsimony and likelihood to coincide but this implies no assumptions that parsimony *requires* that changes be very improbable. In other words, if change is very improbable, for any data set, the most likely hypothesis – represented by a tree and a model of branch transition probabilities – is the one that parsimony would select. From this does not follow that parsimony requires any assumptions about the evolutionary process.

The general idea is that to accept to use the Maximum Likelihood criterion means to reject the possibility of an *a priori* justification of any method of phylogenetic inference, in particular the method of Maximum Parsimony. However, this is an illusion. To use the likelihood theory in phylogenetic inference, that is for composite hypotheses, means to consider any method inevitably linked with some *a priori* assumptions – in the sense of not empirically justifiable – about the character of evolution. For any method it is possible to construct a model of evolution such that the method will give you the most likely tree in that model; this is true for every model of evolution you may choose, even if it is judged to be totally unrealistic.

3.6 Parsimony, Likelihood, and The Principle of Common Cause

The principle of common cause has been formulated by Reichenbach (1956) and Salmon (1975, 1984). It asserts that a set of correlated events that are not related to each other as cause to effect are best explained by postulating a common cause that renders them conditionally probabilistically independent.

Sober (1984a, 1987) suggested some criticisms of this principle. In particular, one of his criticisms involves the global use of this principle:

> I argued that there is no universal principle of scientific rationality that implies that common cause explanations are always better supported than separate cause explanations. Whether this is so depends on the detailed background assumptions involved (Sober, 1987, p.465).

It is our background knowledge that tells us that some explanations of common cause are not plausible: correlations *per se* do not indicate a common cause. Sober (1987) considers as a counterexample to the Reichenbach/Salmon principle of the common cause the method of cladistic parsimony. We have seen that cladistic parsimony does not consider as evidence of common descent the symplesiomorphies but only the synapomorphies. This method therefore implies that mere existence of a correlation – in a total sense – between the traits of two species is not enough to justify a claim of common descent. This claim can find a justification only inside a certain framework of assumptions and inside a specific methodological framework.

Sober (1984a, 1988, 2004) formulated a likelihood justification of cladistic parsimony. He tried to show that the most parsimonious topology of tree inside a framework of specific assumptions is the one of maximum likelihood. Likelihood is often a plausible measure of the explanatory power of a hypothesis and like we have said previously can be useful in deciding on the plausibility of a hypothesis (see above). In fact, parsimony, like every other principle, cannot by itself make one hypothesis more plausible than another.

Let us see how likelihood can be useful in the choice between two phylogenetic hypotheses of overall similarity and cladistic parsimony. In Figure 24 we have two phylogenetic hypotheses for three taxa: robin, sparrow, and lizard. Robins and sparrows have wings, whereas lizards do not.

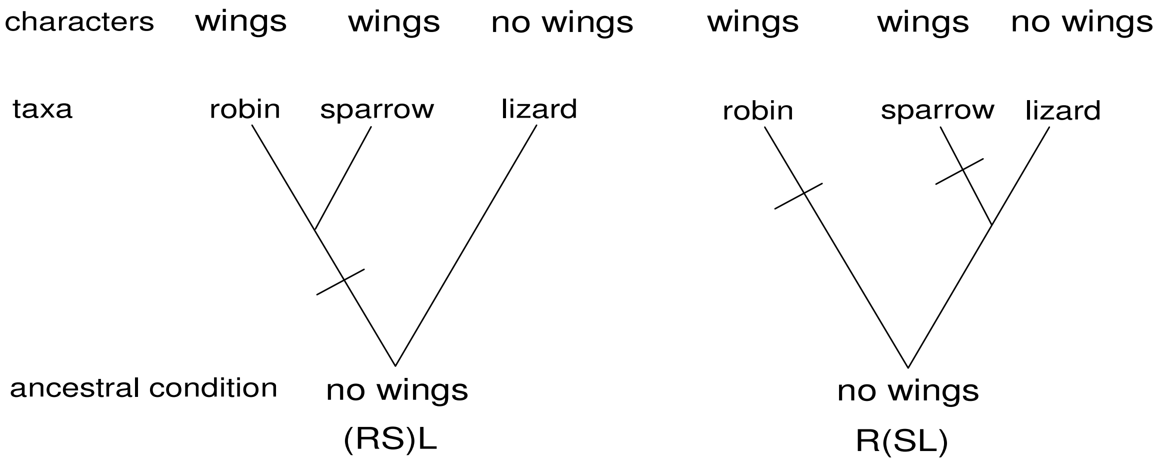


Figure 24 (Sober 2000, modified)

Let us make two evolutionary assumptions. Let us assume that the three taxa share a common ancestor and that this ancestor is without wings. In other words, we are supposing that having wings is the derived (apomorphic) condition and lacking wings is the ancestral (plesiomorphic) state. The (RS)L tree explains this distribution with only one change. The (RS)L tree is consistent with the supposition that robins and sparrows obtained their wings from a common ancestor; this similarity might be a homology. In general, synapomorphies and symplesiomorphies may be both homologies, but *they* *need not be*. When we speak about derived and ancestral characters we indicate a *matching* between two taxa in relation to the character state of some ancestor. But the character of the evolutionary process which links the ancestor with the two matching taxa is not implied by either the two claims of matching (synapomorphies and symplesiomorphies).

For the R(SL) tree at least two changes in the character state are needed in order to explain such a distribution. According to this tree, the similarity between robins and sparrows cannot be a homology, but must be the result of two independent originations. This similarity must involve a homoplasy.

The principle of parsimony says that the tree (RS)L supports the character distribution just mentioned better than R(SL). The reason is that the latter hypothesis requires at least two independent evolutionary changes (homoplasies, given the character state of the ancestor) to explain the data, whereas the former requires only one change (homology, given the character state of the ancestor). In cladistics the principle of parsimony says that *we must choose the tree that minimize assumptions of homoplasy*. Independently of the principle of parsimony, cladism claims that only derived (synapomorphies) and not ancestral (symplesiomorphies) similarities are evidence of propinquity of descent.

The problem is that if we knew that multiple originations were impossible or at least very improbable, then it would be clear why a phylogenetic hypothesis requiring two changes is better than a hypothesis requiring only one. But both these assumptions are problematic. Critics of parsimony (see Sober 1990, 2005) frequently assert that parsimony assumes that evolutionary changes are rare or improbable. The point is that a phylogenetic hypothesis by itself does not tell us whether a given character distribution is to be expected or not: “Genealogies, per se, do not confer probabilities on data”

(Sober, 1987, p.467). Only if we make the assumption that multiple changes are impossible or improbable, then the different hypotheses do make different predictions about the data. In other words, by assuming that multiple changes are improbable, a hypothesis considering only derived characters is better supported: "phylogenetic hypotheses are able to say how probable the observations are only if we append further assumptions about character evolution" (Sober, 1988, p.150). This does not mean that cladistic parsimony requires the homoplasies to be rare, but that this context in which they are rare represents a *sufficient condition* for cladistic parsimony to be right in its judgment that synapomorphies have a significance that symplesiomorphies do not possess.

Consider three taxa A, B, and C. Each character is in two states: 1 and 0 ('1' denote the derived form and '0' the ancestral).

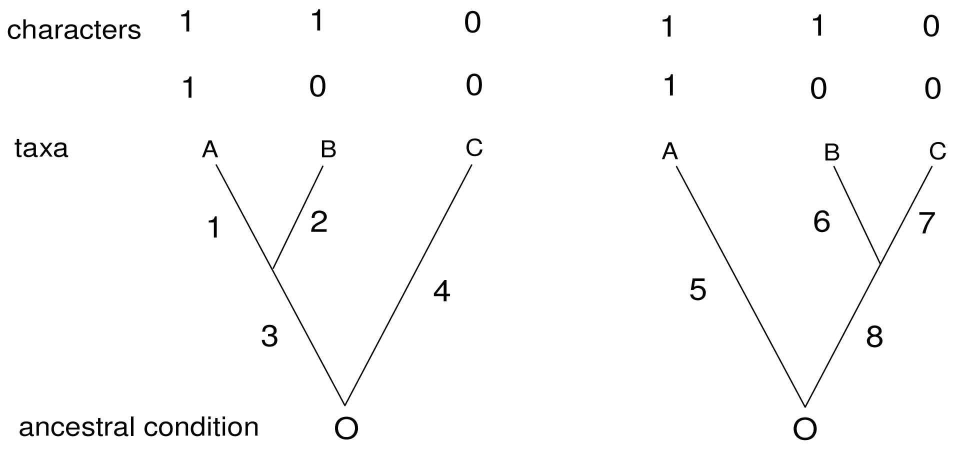


Figure 25 (Sober 1990, modified)

The problem is to infer which two of these three taxa share a common ancestor. We can write the two expressions representing the likelihood of two phylogenetic hypotheses for the character I: P[110/(AB)C]] and P[110/A(BC)]. If we assume that changes on all branches have probability of 0.5, then the two hypotheses will have identical likelihood. If we assume that change is very improbable on branches 1, 2, and 3, but is not improbable on branches 5 and 6, we will obtain that (AB)C is less supported than A(BC).

The idea of using likelihood in order to decide between different hypotheses is that with different assumptions about the branch transition probabilities (probabilities of change) we have different likelihoods for the hypotheses relative on character I. Genealogies do confer probabilities on data only in conjunction with values for the transition probabilities associated with the tree's branches. These probabilities are unknown and they are *nuisance* parameters (see chapter 1, section 6). If we make multiple origination and reversion very improbable rather than impossible, then in this evolutionary context synapomorphies seem to provide very good evidence of relatedness and symplesiomorphies very poor evidence.

Consider now character II. Cladistic parsimony judges such similarities to be meaningless, but if we want to consider them in a non-cladistic framework, the fact that the likelihoods of the two hypotheses are identical depends on the assignment of branch transition probabilities.

Parsimony will determine the best supported hypothesis in the phylogenetic inference only when this hypothesis is the hypothesis with the maximum likelihood.

Specific models of character evolution one adopts make the most parsimonious hypothesis the hypothesis with the maximum likelihood. In sum, the connection between phylogenetic parsimony and likelihood is possible only inside models founded on biological assumptions about evolution. These assumptions on evolution – specific of each model – make sense inside a broader framework of general assumptions on evolution like the molecular clock and the neutral evolution.

3.7. The Method of Maximum Likelihood vs the Method of Maximum Parsimony

Another use of the principle of parsimony consists in preferring the trees with the minimum number of changes needed to explain the characteristics of tip species. While the use of principle of parsimony for the hypothesis of common ancestry is accepted since justified by means of the likelihood, the use of parsimony in selecting trees with the minimum number of changes is controversial, because it seems to require strong metaphysical presuppositions on the nature of the evolutionary processes. In this case there is not simply the acceptance of the more likely common cause explanation of similarities, but we must accept for example that a hypothesis with no evolution is more plausible than a hypothesis postulating one change. Sober (2005) tried to understand this method and its presuppositions by comparing it with the method of maximum likelihood, the method that finds the tree topology that confers the highest probability on the observed characteristics of tip species. This comparison shows that the two method disagrees, but it does not follow from that that in the cases in which they disagree the method of MP should be considered unjustified. However, for a likelihoodist like Sober, this disagreement seems sufficient to suspect the hypothesis of MP in these cases. The illusion is that ML requires probabilistic models supporting a hypothesis and therefore the ML hypothesis has an epistemological value defined by the assumptions of the model, even if they are unrealistic. But, why a topology with the fewest changes should be less plausible and falsifiable than a topology constructed by maximizing some unrealistic characteristics of the evolutionary process? I suggest that we should evaluate them by means of their predictive success by reaching a cross validation in different data sets[3]. For example, consider this following strategy. While

---

[3] See an analysis of the cross validation model for predictive success, Forster forthcoming.

with MP methods we select the trees representing the shortest evolutionary pattern, we could look for cases of short evolutionary patterns for specific characters in specific organisms. When we have found such a pattern we could use it as a model, which in this case it would be a empirical model and not ideal. Then we could test the parsimonious model of evolution by means of its projectibility of the history of other characters in other organisms. This would be a new way of justify the principle of parsimony, independently of the likelihood theory of evidence.

3.8 Methodological Parsimony. On the Principle of Minimum Assumptions

It is worth introducing a third kind of principle of parsimony, which represents the "methodological" parsimony. The idea is expressed by Sober (1988),who argues that the *best* method in phylogenetic inference is that one that needs the minimum number of assumptions to explain the same evidence. we could reformulate this thought by means of the following questions: Is a method that holds one assumption better than a method that holds two or three assumptions? I think the answer should be No, since a good hypothesis has never been judged on the number of concepts we have used to formulate it, but on the possibility to be justifiable, through it explanatory and predictive value. Given that we know that our phylogenetic inferences are founded on unrealistic assumptions, we should ask the same questions and see if we should change our answer: Is a method that holds one unrealistic assumption better than a method that holds two or three unrealistic assumptions? Any method works inside a model of evolution characterized by a different number of unrealistic assumptions. These assumptions are represented by the parameters that in chapter 1 I have referred to as parameters of *nuisance*. Both the Maximum Parsimony and Maximum Likelihood methods (see §5) offer sufficient explanations of the evidence *only* in particular evolutionary contexts, in the sense that they require specific models of evolution. Assumptions make possible our reconstructions of the past, in the sense that they make these reconstructions to become meaningful interpretations of the biological past. The number of unrealistic assumptions makes a difference in the choice of hypotheses only if we consider "truth" the only epistemological value inside them, that is when we believe that the aim of phylogenetic reconstructions consists in being more and less close to the true phylogenetic pattern.

A history which is not mediated by our assumptions is unknowable, therefore the distinction between a phylogeny without ideal models and a phylogeny with ideal models is epistemologically meaningless.

> An event independent of experience, 'objective' in the sense of being untouched by thought or judgment, would be unknowable; it would be neither fact nor true nor false, but a nonentity. And, in so far as history is a world of facts (which will scarcely be denied), it is a world of ideas, and a world which is true or false according to the degree of its coherence. The distinction between history as it happened (the course of events) and history as it is thought, the distinction between history itself and merely experienced history, must go; it is not merely false. It is meaningless (Oakeshott 1933, p. 93).

The idea of unrealistic assumptions and nuisance parameters is problematic only inside an epistemological framework in which we think that our historical hypotheses should represent the reconstructions of the past as it is, under the assumptions that the knowledge of something represents the knowledge of something as it is. However, whether we hold a neo-Kantian framework, which I would call "illuminated" realism, our representations of history have a meaning which is different on the basis of the assumptions – even if unrealistic! – we make. Our representations can even be projectible. For example, geneticists know that most of the assumptions they use in a model are false, but they continue in using it, because a false model can make accurate predictions. Wimsatt (1994, pp. 498-501) claimed that some of the theories in genetics are inconsistent with the evidence and that the best (in the sense of projectibility) scientific theories are idealizations, deliberate simplifications that are, strictly speaking, false, but are heuristically useful for organizing and structuring the knowledge. Phylogenetic hypotheses are possible inside ideal models of assumptions, and they give us a probable knowledge of the past. Inside the probabilistic "mode", the ontology is given by posterior probabilities, likelihoods and prior probabilities; the epistemology by the theories of Bayesianism and Likelihoodism; the methodology by the calculus of average likelihoods and maximum likelihoods of phylogenetic hypotheses. This mode does not represent a mere construction in the sense that it is independent of the evidence, the world, since any kind of probability, even the prior probability, it is influenced by the reality trough the whole knowledge we have of it. Different modes and models we judge on the basis of their coherence represent the conceptual organization of an evidence, not a structure of the mind.

In sum, our way of experience history is given by probabilistic models and the new ontology we have created for our representations of history is an ontology of probabilities.

> We call a proposition true, not because it agrees with a fixed reality beyond all thought and all possibility of thought, but because it is verified in the process of thought and leads to new and fruitful consequences. Its real justification is the effect, which it produces in the tendency toward progressive unification. Each hypothesis of knowledge has its justification merely with reference to this fundamental task; it is valid to the degree that it succeeds in intellectually organizing and harmoniously shaping the originally isolated sensuous data. (Cassirer 1910, p.318)

4. Assumptions

We have seen that there are general background assumptions and assumptions characteristic of models. One of the most important background assumption is that one concerning the general character of evolution, expressed by means of the equilibrium between the two main forces shaping the biological history: genetic drift and natural selection. The first one acts by increasing genetic distance between populations and reducing genetic variability inside the populations and it said to lead the neutral part of evolution. By using a neutral context of evolution (without selection), geneticists can fix *constant* mutational rates for every branch of the tree and so they can set a molecular clock, either between sequences of different species or between populations within species. But a molecular clock is reliable only if *two* assumptions are true: the genetic similarities of organisms are function of their relatedness and they are directly proportional to the recency of their divergence from a common ancestor. These assumptions are true when mutations have uniform rates in time and they are neither be culled nor favoured by natural selection. Basically, evolution must not be generally convergent. Two are the consequences of natural selection on the reconstruction of phylogeny: culling tends to conserve a gene in its pristine state, so that the common ancestor seems much more recent than it really is; on the other hands, selection for particular mutations accelerates the change, making divergence seem more ancient. Yet, another requirement for a reliable molecular clock, is that the lineages should not have mixed significantly after they have diverged (no gene flow, that is migrations). Otherwise they will seem to have split more recently than they did. Nevertheless, these assumptions are not generally empirically justifiable and they are not plausible in many particular cases. For example, there is substantial evidence that the evolution of mitochondrial DNA is not neutral and it has also been shown that mitochondria do not obey a molecular clock.

It stands to reason that our reconstructions are ideal representations made in ideal evolutionary contexts, without natural selection and gene flow.

Besides these background assumptions, there are other assumptions which constitute the probabilistic models. Some of these assumptions for models in molecular evolution are:

- All characters change in an independent way.
- Any character changes with the same probability.
- The evolutionary rate is uniform.

These assumptions represent constraints used to reduce the number of adjustable parameters of a model. For example, consider a tree in which two characters A and B have changed in the two tip species X and Y.

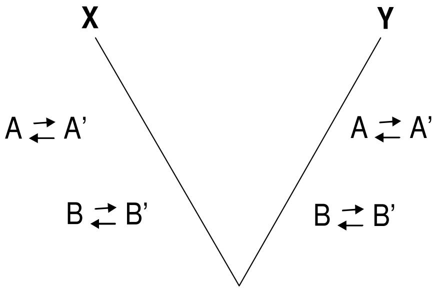


Figure 26

If we assume that the any character changes with the same probability, we have that $p_A = p_B$

Then we can assume that reversions are very improbable, and we will have that $p_{A'} = p_{B'} \sim 0$. In this way it is easy to understand that we have to give a value to only one parameter instead of 4.

Models are simpler and more complex depending on the number of adjustable parameters, therefore depending on the number of idealized assumptions they require. As we will see in the next section about the principle of parsimony, an alternative strategy to the ideal models would be possible. For example, instead of assuming that characters change with the same probability and that the evolutionary rates are uniform trough the time, we could imagine to study the variation of a character in an organism and then use this organism as a model for the evolutionary rates of other organisms, empirically testing the hypothesis of a specific evolutionary pattern.

5. Genealogical Hypotheses and Prediction

5.1 The Problem with Non-Probabilistic Evolutionary Models

Consider the evolution of a single character as it is passed from parent to offspring, where the character has states 0 or 1.There are 4 possible histories shown in the table on the right. Each history refers to a hypothesis on the evolution of the character. These maximally specific hypotheses are labelled h_1, h_2, h_3, and h_4, as shown in the table. Such a hypothesis makes specific predictions about the character state in any generation. All other (non-probabilistic) hypotheses can be constructed as disjunctions of these. We sometimes refer to such disjunctive hypotheses as *models*.

	Parent	Offspring
h_1	0	0
h_2	0	1
h_3	1	0
h_4	1	1

Suppose that we are told the state of the character of the parent. How could we use this evidence to learn about the character state of the offspring? If we can’t make prediction using this information, we certainly will not succeed without it. Assume that we know that the disjunction ($h_1 \vee h_2 \vee h_3 \vee h_4$) is true prior to seeing the evidence, but that's all we know. If the parent is in the 'derived' state 1, we learn that h_1 and h_2 are false, so now we know that ($h_3 \vee h_4$) is true. The problem is that we still know nothing about the offspring character state, or at least nothing new. For h_3 predicts that the offspring character will be 0, while h_4 predicts it will be 1, so ($h_3 \vee h_4$) predicts that the offspring state be 0 or 1, which is something we already knew.

The same point applies to the problem of postdicting the state of the parent from the state of the offspring—*we can't predict the past from the present if we only assume that some history is true*.

The point applies very generally. We could consider an arbitrary number of generation, and arbitrary number of characters possessed by each individual, and an arbitrary number of individuals. More concretely, suppose that an organism evolves asexually for 100 generations, and we observed that in the first 99 generations the character state is 0. No amount of data about the past will predict the future *by itself*, just as no amount of data about the present will predict the past, *by itself*. You need to postulate some hypothesis before you can make predictions. If there is no model, then there is no inference (Sober 1988).

For instance, consider a model such as *M*: If a change from 0 to 1 (a mutation) occurs, then it will never reverse. If we consider only two generations, then $M = (h_1 \vee h_2 \vee h_4)$. This hypothesis enables us to predict that any offspring will be in state 1 from the fact that the parent is in state 1. Let *e* be the assertion that the parent is in state 1. The datum *e* implies that hypotheses h_1 and h_2 are false. In other words, the conjunction of *M* with *e* is logically equivalent to h_4. Let *p* be the prediction that the offspring is in state 1. Clearly h_4 implies *p*, so we have demonstrated that the conjunction (*M* & *e*) logically entails p. The example illustrates the general manner in which data is combined with a model in order to make predictions. A model is a disjunctive hypothesis, which does not entail a precise history, but when combined with data about the character states in some generations, it can make an exact prediction about the state of the character in other generations.

It would be a mistake to conclude that the model by itself makes no observable predictions. For if (M & e) entails p, then M entails (if e then p). Thus, the previous story about how M makes the prediction p could be retold in the following way: M entails (if e then p), and so if we observe e then we can deduce p by *modus ponens*.

Hypothetico-deductivism (H-D) is a normative view about how science should work, and does work in many cases. It makes a distinction between the context of discovery, the hypothetico part, from the context of justification, the deductive part. In the context of the discovery, or invention, or construction, scientific theories, models, and hypotheses, can be freely postulated, or hypothesized. There is no logic of discovery on this view. It is often assumed tacitly that the space of hypotheses should be unconstrained for if we were to restrict the space of hypotheses *a priori*; that is, without looking at the evidence, then we would be introducing an undesirable kind of bias or subjectivity into science. Although this is not one of the core assumption of the H-D view.

In the context of justification, the hypotheses should be tested by deducing observable predictions from them, and then comparing these predictions with what is actually observed. On the H-D view, the relationship between theory and evidence is deductive, not in the direction from evidence to theory, but from theory to evidence. It doesn't specify exactly how theories or hypotheses should be evaluated from or judged in terms the observed truth or falsity of predictions.

Falsificationism is an H-D view that goes a step further. It claims that any hypothesis that is falsified by the evidence should be rejected since it is proven to be false. Those that are unfalsified may also be false, but at least they have a chance of being true. But which of the many unfalsified hypotheses should we favor? Falsificationism adds the rule that more falsifiable hypotheses should be favored over less falsifiable hypotheses. Falsifiability is a purely logical property that has nothing to do with what is actually observed. It is an non-empirical virtue that may be used to break the tie amongst hypotheses that fit the data equally well.

How does falsificationism fare in this example? The set of possible hypotheses is $\{h_1, h_2, h_3, h_4\}$ together with any disjunctions of these. But a disjunction is always less falsifiable that any of its disjuncts. That is, $\{h_1, h_2, h_3, h_4\}$ is the set of most falsifiable hypotheses, and they are all equally falsifiable. Consider the evidence that the parent is in state 0. That tells us that h_3 and h_4 are false, which leaves $\{h_1, h_2\}$ as the set of unfalsified hypotheses. But there is no way of favoring one over the other because both are equally falsifiable. The hypothesis h_1 predicts that no mutation will occur, while h_2 predicts that the final character state will change to 1. So, if we treat all hypotheses as viable competitors, and apply only the falsification and falsifiability criteria, then we are unable to favor h_4 over h_3, or *one prediction over another*. This point says more than no-model no-inference slogan. It says that if all models are considered, then the most falsifiable unfalsified hypotheses will make precise predictions, but *the predictions will be different*. The methodology of falsificationism applied to non-probabilistic hypothesis fails to provide any rationale for favoring one prediction over the other.

The same analysis carries over to the case in which we observe the character state for 99 generations, or to the problem of predicting past character states from the present state. We are still left with equally falsifiable unfalsified hypotheses that make different prediction.

Bayesianism enriches the falsificationist framework by assigning prior probabilities to each maximally specific hypothesis, that is, to the *h*'s, which specify the exact evolutionary histories. If there are *N* possible histories, then there are *N* maximally specific hypotheses h_i for $i = 1, 2, 3, \ldots, N$. Learning is then characterized in terms of a change from the prior distribution of probabilities over the h_i to a posterior distribution of probabilities. The learning algorithm determines the posterior probability of h_i from the prior probability of h_i *given* the evidence, which is denoted by *e*. If we denote the prior probability function by *P* and the posterior probability function by *P′*, then for any h_i, $P'(h_i) = P(h_i \mid e)$. What's learned therefore depends on the differences between the posterior and prior probability distributions over the hypothesis space. This can be represented by the differences of the form $P(h_i \mid e) - P(h_i)$, which is described as the degree to which *e confirms* h_i. When the difference is negative, then we say that *e* disconfirms h_i.

Consider the general case in which the organism evolves for 100 generations, and *e* specifies the character values on the first 99 days. How do Bayesians treat this example? Well, there are Bayesians, and there are Bayesians. Subjective Bayesians think of an arbitrary probability distribution of the *h*'s as representing a state of belief, or partial belief if none of the *h*'s have probability one. Different prior distributions represent the states of belief of different people, which is why is called Subjective Bayesianism. Learning is the change between two states of belief determined by the belief-updating rule $P'(h) = P(h \mid e)$. The most apparent problem with Subjective Bayesianism is that what's learned depends on the *initial belief state*. In itself, this does not seem to be unreasonable. But when we ask about why it is rational to adopt one initial state of belief over another, the answer is “seems right to me” or “no reason”.

Objective Bayesian are different in this respect. They argue that in this example, when the initial state of belief is formed prior to seeing any evidence, the prior distribution should represent a state of complete ignorance, and this is captured uniquely by the distribution that assigns equal prior probabilities to all of the 2^N possible histories. We have two possible histories because each of the N generations can be in state 1 or 0. Bayesians refer to this as the *ignorance prior*.

Posterior probabilities are calculated using Bayes theorem, which is a theorem of probability theory:

$$P'(h) = P(h \mid e) = P(h)P(e \mid h)/P(e).$$

$P(e \mid h)$ is the *likelihood* of *h* relative to *e* and we have seen that it is defined as the probability of the evidence *e* given *h*. The difference between the probability of *h* and the likelihood of *h* involves a reversal of the relationship between theory and evidence familiar in the H-D view of science. In our example, for any particular h_i, *e* is either deduced from h_i, or it falsifies h_i. If *e* is deducible from h_i, then the likelihood of h_i takes on its maximum possible value of 1. If h_i is falsified by *e*, then its likelihood has the smallest possible value, 0, in which case Bayes theorem proves that the posterior probability must be 0. So all Bayesians (objective and subjective) agree with their falsificationist cousins that refuted hypotheses should be rejected in the sense that their posterior probabilities diminish to zero. If h_i entails *e*, we have that the likelihood of h_i is equal 1 and then the posterior probability of h_i is

$$P'(h_i) = P(h_i \mid e) = P(h_i)/P(e),$$

Hypotheses that survive falsification all entail *e*, so their probabilities have increased by the same multiplicative constant factor $1/P(e)$. This means that for any two histories h_i and h_j that are compatible with the evidence *e*, the ratio of their posterior probabilities is equal to the ratio of their prior probabilities. Therefore there is nothing we have learned about the relative plausibility of the possible histories except that some are falsified by the evidence. We see that Bayesianism adds nothing to a falsificationist methodology beyond the assignment of a prior probability to every possible history.

As previously mentioned, objective Bayesians begin with an ignorance prior in which all (maximally specific) hypotheses are equally probable. In this case, any hypothesis about the future is probabilistically independent of any statement about the past, and vice versa, in the sense that any hypothesis about the past is probabilistically independent of any statement about the present. In this last case, for example, if *e* is a statement about the present character state, for example, and *H* is *any* hypothesis about the character states in past generations, then $P(H\,\&\,e) = P(H)P(e)$ [4], and so $P(H \mid e) = P(H)$. The reason is that an ignorance prior is constructed by assigning a probability of ½ to a character state *e* in each generation, and then multiplying the probabilities together, so that each possible history has a probability of $1/2^N$. So, each history has the same probability, and the probabilities add to 1 because there are 2^N histories. If we hypothesize no correlation between the past and the present then there is nothing that can be learned about the past from the present. In the machine learning literature, this is known as the *no-free-lunch theorem.* The lesson usually taken from this theorem is that if there is no bias built into the prior probabilities, then there is no learning.

---

[4] Given two events a and b, if they are independent, then $P(a\,\&\,b) = P(a)P(b)$

If, instead, we begin with a non-uniform prior, as subjective Bayesianism allows, then *H* and *e* can be probabilistically dependent even though they are logically independent (neither entails the other). We end up “learning” something about the past from the present in the sense that $P(H \mid e) \neq P(H)$. But what we learn is completely accounted for in terms of the falsification of histories in the disjunction *H* that have been falsified by *e*.

In sum, Bayesianism agrees with Falsificationist principles to the extent that any falsified history has probability zero. The difference is that for a Bayesian, the corroboration, or confirmation, of surviving hypothesis is not determined by their falsifiability, but their prior probability. This is an important difference because these forces act in opposite directions. If hypothesis *A* entails hypothesis *B*, then the probability of *A* is less than or equal to the probability of *B*. So probability will tend to favor *B* over *A*. But if *A* entails *B*, *A* is more falsifiable than *B*, so the falsifiability criterion favors *A* over *B*. Any tautology has probability one. High probability is easy to achieve if there is no requirement that the hypothesis is informative.

Likelihoodism offers the possibility of thinking whether hypotheses imply what character distributions *will probably* occur instead of thinking whether hypotheses imply what character distributions occur. For by holding the hypothetico-deductive approach we are forced to analyse the possibility of hypotheses to imply what character distribution *will* occur. In this sense the likelihood approach provides a probabilistic analogue of the concept of strong falsification. If two hypotheses each imply a probability for a distribution, the one conferring the higher probability is better supported by the evidence. However, as I pointed out, hypotheses either do not imply character distributions and they do not imply how they are *probably* distributed. Only a tree that is conjoined with a model of specific branch transition distribution does imply a *probability* for each possible character distribution (see §3.6). The probabilistic mode of evidence offers a conceptual framework in which we can formulate phylogenetic hypotheses. However, there is still an empirical issue, which concerns the values we decide to give to the parameters of our models, like for example the values of the transition probability; in a phylogenetic inference not only we are ignorant of the values of these probabilities.

5.1.1 Parsimony in Non-Probabilistic Models

In any evolutionary history, we can count the number of times that the character state changes, says from 0 to 1. For instance, if there is 5 generations, the histories 00000 and 11111 involve no changes. Let us define the *parsimony* of a hypothesis by the number of state changes that it postulates. The hypotheses that postulate histories 00000 or 11111 are *maximally parsimonious*. On the other hand, if the hypothesis postulates 0111, it is less parsimonious, but more parsimonious than on that postulates that 01011, which involves 3 changes. Parsimony is not the same as falsifiability because all specific histories are equally falsifiable, but not equally parsimonious.

We can also introduce evolutionary models as the disjunction of all histories that have a fixed degree of parsimony, and it is then natural to define the parsimony of model in terms of the parsimony of the disjuncts. The most parsimonious model of this kind is the one that asserts that the true history is one of two histories, 00000 or 11111 in the example with 5-generations. If *e* is the observation that the character state in generation 5 is 1, then (*M* & *e*) entails that the history is 11111, which makes very precise predictions about the past. On the other hand, the model that postulates one change is the disjunction of ten possible histories: 00000, 00001, 00011, 00111, 01111, 11111, 11110, 11100, 11000, and 10000. If *e* is the observation that the character state in generation 5 is 1, then in this case (*M* & *e*) entails that the history is 00001, 00011, 00111, 01111, or 11111, which makes less precise predictions about the past. There appears to be a connection between the parsimony and predictive power of models, which might explain the methodological appeal of parsimony in phylogenetic inference.

Another way of describing the model is in terms of adjustable parameters; according to the hypothesis, a specific history is characterized by its initial state and the generation at which the character changes from 0 to 1. If the final state is observed to be 1 then the initial state must have been 0 because the model postulates a single change. And if the final state is observed to be 0 then the initial state must have been 1, according to the model. Thus, the initial state parameter is determined by observing the final state. The transition time, however, is undetermined by observing the final state, so the complete set of parameter is left underdetermined by the character states of present-day organisms, and the model is said to be *non-identifiable*.

A more standard curve fitting example of the determination of parameters from data is depicted in Figure 27. In the case of a model such as LIN, $y = a + bx$, in which there are two adjustable parameters, there have to be at least two data points to exactly determine their values. If there are fewer data points, then the parameter values are undetermined by the data. In the usual case, in which there are more data than adjustable parameters, then the values determined by a subset of the data are then tested by the remaining data. In this case, the model can display a genuine kind of predictive success in which part of the data can be used to successfully predict features of the remaining data.

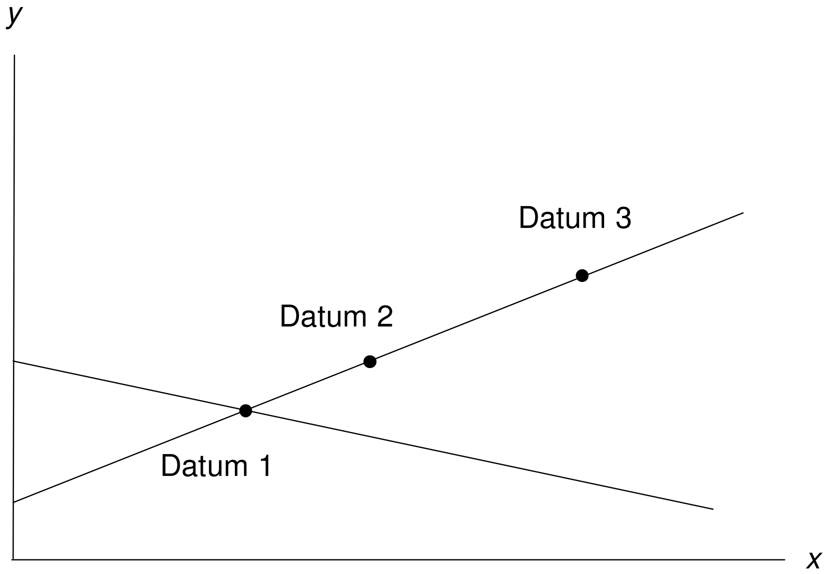


Figure 27: Curve fitting example. The adjustable parameters in the model $y = a + bx$ are underdetermined if the data consists only of Datum 1 because there are many lines that fit one data point. The parameters are exactly determined if there are two data points. If the data consists of three data points that are consistent with the model, then the parameters are overdetermined. Two points determine the line, and the third point tests the model.

If the number of data is less than the number of adjustable parameters, then not all the parameter values are determined by the data. But it does not follow that the parameter values are determined if there are more data than adjustable parameters. For example, in the Figure, the three data points could be on top of one another if they had the same value of *x*. In that case, the parameter values would be underdetermined by the data even though there are more data than adjustable parameters. It is not the number of data that counts, but whether they contain sufficient information. This seems to be the same problem in phylogenetic inference. Many facts about the past may be inaccessible to us no matter how much we know about the present.

It is puzzling that the most parsimonious model postulates no change, and therefore no evolution. If we only observe the character state of present generation, and define the best explanation of our observations to be given by the most parsimonious model compatible with those observations, then it appears that the best explanation is going that there is no evolution. How can it be that evolution is the best explanation of observed facts about the character states of present-day organisms? Is it because science does not define 'best' in terms of parsimony, as defined, or is it because the no-evolution model is not compatible with the evidence? Let us investigate the second possibility by considering richer kinds of data about present-day organisms.

Suppose that there are three character states 0, 1, and 2. The observation of 'derived' character states cannot provide evidence for evolution without independent evidence that the character was in a 'primitive' state in past generations. But this is exactly what is at issue. Likewise, if we consider more than one character, the most parsimonious model is still one that postulates that there has been no change in any of the characters. The only kind of evidence that could possibly provide evidence for evolutionary change must refer to the similarities and dissimilarities amongst different organisms. The challenge is to find a single clear-cut case of this kind.

Let us try the simplest kind of example first. Suppose we compare one organism sampled from one population, and a second organism sampled from a second population. If the two populations can interbreed then it would not be surprising that they have common ancestry.

So, suppose that the two populations cannot interbreed at the present time. This means that they must be dissimilar in same ways, and so we need to assume that some characters are dissimilar and some characters are similar. Can the similarities provide evidence of a common ancestry? The question is whether the evidence before us is compatible with the maximally parsimonious model, which says that the organism ascended from different ancestors with no changes in any of their current character states. This model is more parsimonious than any model that postulates a common ancestry because any common ancestry model would have to postulate some changes in order to account for the differences that currently exist.

6. On the Principle of Total Evidence (TE) and Correlation in Phylogenetic Systematics

First, let us remind what the principle of TE is. Good (1983:178-180) used a mathematical theorem to demonstrate that the principle of total evidence follows from the principle of rationality. The principle of total evidence is an important maxim in phylogenetic systematics, because alternatives to using all available evidence must be carefully considered. For example, in order to exclude logically consistent evidence, the investigator must argue that "the premises constitute either the total evidence *e* available at the time or else a part of *e* which supports the conclusion to the same extent as does *e*" (Hempel, 1965:64). Thus, even less severe weighting, as well as partitioning evidence, requires justification. Furthermore, including all relevant evidence can be seen as a harmless activity, unless one is prepared to argue a priori that certain evidence will confound the analysis and must therefore be eliminated. The assumption of independence of evidence exists in all empirical sciences. Ideally in phylogenetic systematics, each synapomorphy in the character matrix is assumed to count as a separate piece of evidence, has the potential to confirm, or disconfirm, taxonomic relationships independent of all other synapomorphies considered. This is important, because choice of a phylogenetic proposition is based on the largest number of synapomorphies that can be hypothesized as homologues. In actual practice, independence of evidence can be viewed as a weighting criterion - the phylogenetic hypothesis confirmed with the largest number of synapomorphies with the greatest likelihood of independence is preferred.

A recent paper by Fitzhugh (2006) analyses the principle of TE in phylogenetic inference. He claims that this principle is needed in reconstructing phylogeny, because it is needed in all non-deductive inferences; however, it is violated in phylogenetic inference, since partitioned data sets from the TE are independent[5] and therefore cannot be demonstrated to be evidentially irrelevant to one another. Since partitioned data sets are independent of each other then the choice of separate pieces of evidence always represents a violation of the principle of total evidence. However, first he fails in understanding the difference between the context of discovery from the context of justification. For, he considers the principle of TE needed to be true in order to use partitioned data sets in phylogenetic inference, without considering the possibility of a justification of the hypotheses explaining partitioned data sets. Yet, he does not justify the belief in such an independency between data sets and I argue that there is no reason to assume that data sets are independent. The covariation of character states has never been taken seriously in consideration in phylogenetic systematics. Covariation, measured by statistical correlation between characters, represents the evidence that should be explained by phylogenetic hypotheses. Figure 28 describes the frequencies with which a set of characters exhibits the four possible character distributions. In the case a) the characters have no association, in the case b) they have a positive association.

---

[5] Two data sets are independent if there is zero association between them.

a)

	25%	25%	25%	25%
A	1	1	0	0
B	1	0	1	0

b)

	25%	15%	35%	25%
A	1	1	0	0
B	1	0	1	0

Figure 28. Two character distributions for the two species A and B. In a) there is zero association; in b) there is a positive association.

Why correlation between characters should be better than many observations of single matchings? Since just in this way we could understand the rules leading the evolution of the characters and have more empirical information about the process. The knowledge of correlations represents a very important information in science in the unification of concepts. Usually we start looking for a significant agreement of independent tasks on postulated quantities, where this agreement can be expressed by an empirical *overdetermination* of some postulated quantities, constants or concepts. Secondly, the agreement must be *statistically significant*, so that it is implausible to explain the agreement on the basis of pure chance. Once we have reached this statistical significance, we can postulate a law unifying the concepts.

> The "objectivity" of pure concepts and truths is accordingly put on a plane with that of physical things. Nevertheless, the difference between them is sharply revealed again, when we recall that in the sphere of logics and mathematics we cannot reach absolute "objects", but always only relative objects. Not *number*, but

> rather the numbers constitute a true “entity”. Here the individual gains its meaning and content only from the whole; - but this whole cannot be presented like a quiescent object of perception, but, in order to be truly surveyed, must be grasped and determined in the law of its construction. In order to comprehend the number series *as a series* and thus to penetrate into its systematic nature, there is needed not merely a single apperceptive act (such as is considered sufficient of the perception of a particular thing of sense), but always a manifold of such acts, which reciprocally condition each other. (*Ivi*, p. 317)

Matchings *per se* are no evidence of phylogenetic relationships and independent hypotheses are perfectly sufficient to explain them. The hypothesis of a common ancestry is needed only when we can demonstrate that these matchings have an additional significance, given by the positive statistical correlation between character states. This is exactly what M. Forster (1988) has pointed out about common cause explanation. He claims that common cause hypotheses explain only correlation and not series of co-occurrence of events, that is what Cassirer calls the law we must grasp under the series of “apperceptive” acts. We can observe an infinite number of matchings, but if we do not understand the law that generates this evidence, they rest only meaningless experiences.

In sum, the causal question we should ask about two species is not:

Why is character X in state 1 in species A and in state 0 in species B? Why is character Y in state 1 in species A and in state 1 in species B?, etc.

But: Why are the states of characters X, Y, etc. associated in a certain way between two or more species? What is the law explaining this association?

Consider the Aristotelian's example of the market in *Physics*, Book II, chapter 5. There are two men meeting at the market. The meeting can be considered fortuitous and explained by the hypothesis that the two men go to the market for independent reasons. The explanation of the meeting is given by explaining separately why the first man goes to the market and why the second goes to the market. But if I observe that once a week, week after week the two men keep on meeting then we can have the suspect that another explanation would be possible, since perhaps both men have a common reason in choosing when they go to the market, for example they can be both in love with the fishmonger's daughter, who is at the market once a week. The meetings *per se* do not require a common cause explanation, since they can perfectly explain by saying that they go to see *a* girl; in order to formulate a common cause explanation we need to know or to assume that the two men go to the market for the *same* girl. What a common cause explanation can explain is not mere matchings, but statistical correlations. Given two events a and b, we have positive statistical correlation between them if the co-occurrence of the two events a and b is *greater than* the occurrence of the two events separately. Therefore what a common ancestor hypothesis can explain is not the co-occurrence (matchings) of a state for two characters in two species, but the statistical correlation between the character states. If we have no information about the "significance" of the co-occurrence, then the common ancestor hypothesis is only *sufficient* in explaining the matchings but it is *not needed*.

At this point we could ask: What is the evidence of evolution? Whether the common origin hypothesis is needed to explain evidence in a historical way in the sense of giving a historical sense to the present evidence, then our evidence should be given by correlations between character states and not matchings.

Characters can be historically correlated because they are the result of the same biological process or mechanism, independently whether they share a common ancestor. There are two kinds of histories we should consider: the history 1 made by the evolutionary relationships of individuals inside each species and the history 2 of the phylogenetic relationships among different species. The first history refers to the evolutionary process that shapes single species from the past to the present long each branch; the second to the phylogenetic relationships among different species and their ancestors from the past to the present.

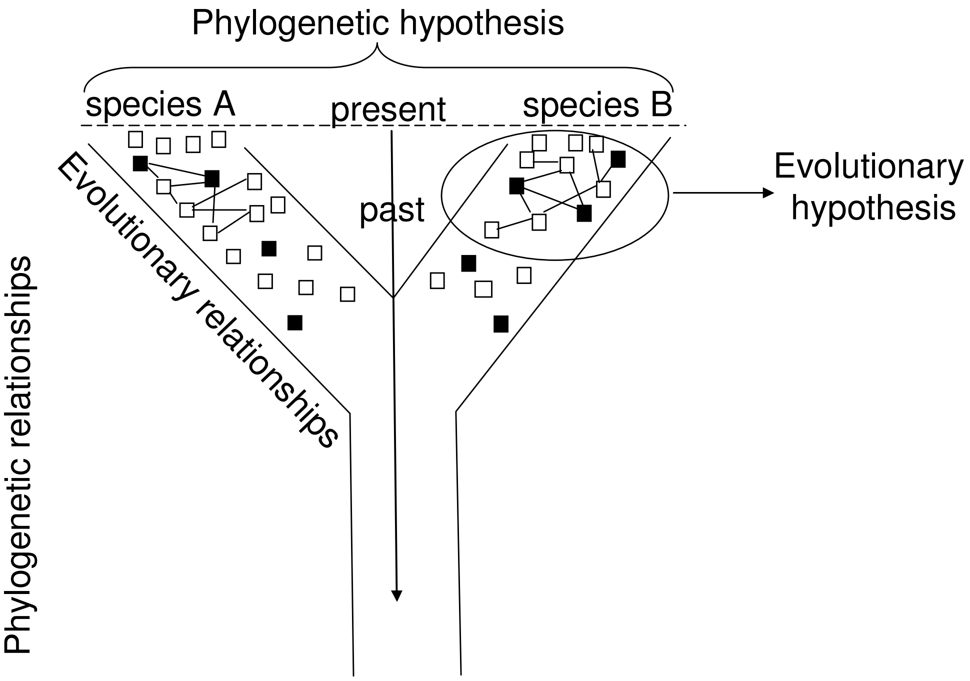


Figure 29. History1 and History2. (modified from Hennig 1966: fig. 6)

The historical question regarding the first history is a causal question of the process producing some characteristics of the species. The historical question regarding the second history is the causal question answered by a hypothesis – the phylogenetic hypothesis – explaining the similarities and differences among present species, by assuming a certain history 1 and a certain common ancestor, whether we want to choose the common ancestry hypothesis as the most plausible. Statistical correlations between characters will give part of what I have called “history 1”, which is now determined by means of ideal constraints.

7. Representations in Phylogenetic Systematics

In the previous chapter, I have presented four epistemological possibilities of understanding the past in historical disciplines and I have suggested the idea of considering historical hypotheses as historical representations of the evidence. Any tree is an image of a history of species, constructed on a certain evidence, given a specific model of evolution. Each model with its adjustable parameters is characterized by the assumptions on the evolutionary process and it makes possible a representation of history, that is, a topology. I have called the models “ideal”, not only because they are based on ideal assumptions of the evolutionary process, but also to stress the analogy with Weber’s ideal types. Like every historical hypothesis, the phylogenetic hypothesis is a possible scenario made up with a set of causes, which are finally *conjunctively sufficient* to give to the evidence a historical meaning.

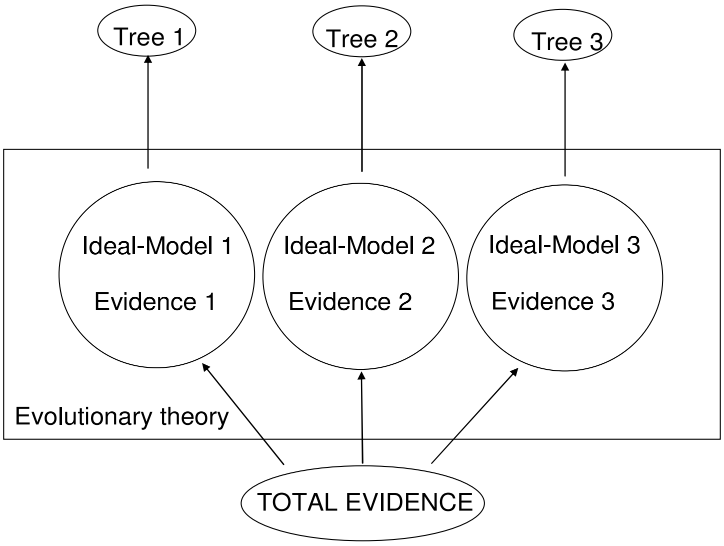


Figure 30. The Structure of Phylogenetic Inference

These models can compete on the basis of coherence and predictive accuracy but not on the basis of truth. Depending on the number of the adjustable parameters they are more and less complex, but all of them are *unrealistic*, since unrealistic are their assumptions on the process. We have seen in chapter 1 that more complex models always have higher likelihood than simpler; the question is whether the likelihood of a more complex model is *sufficiently* greater than the likelihood of a simpler model to justify rejecting the simpler model. Of course, "sufficiently" is a vague term, since we could ask how much greater this likelihood should be in order to reject the simpler model. Like it happens in different issues in statistics, the decisions are often made *a priori*, by establishing a certain boundary-value over which the simpler model can be "reasonably" rejected.

Our beliefs in phylogenetic inferences are founded on several epistemological assumptions:

- The evolutionary theory is true (the accepted framework).
- A genealogical classification exists and shows genealogical relationships among species.
- The hypothesis of common ancestry is the best explanation for similarities.
- The phylogenetic reasoning is essentially a reasoning in accordance with the principles of probability.
- Each biological history is only one of the sufficient explanations of the evidence and it is possible *only* inside an ideal-model which assumes a specific evolutionary process.

In a naïve realist framework, we are worried about the number of assumptions and the unrealistic character of them. We have seen that the hypothesis of common ancestry is the hypothesis of the minimum cladistic parsimony and that the principle of parsimony is not a very reliable principle. In a naïve realist epistemology we should be worried about it because the only knowledge we want is the knowledge of the world in itself, how it is. However, inside a "illuminated" realist epistemology, principles as the principle of parsimony and assumptions like an evolution with no selection are conceptual tools needed to grasp a meaning of the past. The knowledge of the past is not about reconstructing the true pattern, but it is about grasping the meaning that different phylogenetic hypotheses inside their models can give us. Our representations of the past will be meaningful, even if they are based on assumptions considered as unrealistic. In an "illuminated" realism, there are meaningful idealizations and meaningless idealizations: the first ones should be eliminated and the second should be kept in science. Phylogenetic inference is not a recalling of evolutionary events, it is a historical reasoning on the biological evidence by means of ideal models on the evolutionary processes and an interpretation of the presuppositions of these models. The interpretation of the conceptual structures needed for reconstructing phylogeny is a guide for the biological research on the evolutionary processes, which would be blind otherwise.

I like to conclude my dissertation with Oakeshott's words on the nature of the historical inquiry.

> My view is, then, that the theory of knowledge at the back of the notion that history begins with the collection of material and that the data in history are 'isolated facts', is erroneous, and that the notion itself is preposterous. We know nothing of a course of historical events apart from some system of postulates; and it is the first business of anyone who is considering the character of history to discover the nature of these postulates. And further, no line can be drawn between what is presupposed and what is known. What is known is always in terms of what is presupposed (Oakeshott 1933, p.98).

BIBLIOGRAPHY

Akaike, H. (1973). “Information Theory and an Extension of the Maximum Likelihood”. In *Second International Symposium on Information Theory* (B. Petrov & F. Csaki, eds.). Budapest: Akadémiai Kiadó.

Andreasen, R.O. (2000). “Biological Reality or Social Construct?”, *Philosophy of Science*, 67, Supplement. Proceedings of the 1998 Biennal Meetings of the Philosophy of Science Association. Part II: Symposia Papers: S653-S666.

Barbujani, G. and Belle, E.M.S. (2006). “Genomic Boundaries Between Human Populations”, *Human Heredity* 61: 15-21.

Baum, D.A. and Shaw, K.L. (1995). “Genealogical Perspectives on the Species Problem”. In *Experimental and Molecular Approaches to Plant Biosystematics* (P.C. Hock & A.G. Stephenson, eds). Monographs in Systematic Botany from the Missouri Botanical Garden, 53: 289-303.

Bloch, M. (1953). *The Historian’s Craft*. Alfred A. Knopf: New York.

Becker, C. (1910). “Detachment and the Writing of History”, *Atlantic Monthly*, CVI, October 1910: 524-536.

Boniolo, G. (2007). *On Scientific Representations. From Kant to a New Philosophy of Science*. New York: Palgrave Macmillan.

Brogaard, B. (2004). “Species as Individuals”, *Biology and Philosophy* 19: 223-242.

Burchard, E.G. et al. (2003). “The Importance of Race and Ethnic Background in Biomedical Research and Clinical Practice”, *The New England Journal of Medicine* 348, 12: 1170-1175.

Cain, A.J. (1958). “Logic and Memory in Linnaeus’ System of Taxonomy”, *Proceedings of the Linnean Society of London* 169: 144-163.

Carr, E.H. (1961). *What is History?* London: Macmillan.

Cassirer, E. (1910). *Substanzbegriff und Funktionbegriff*. Engl. Trans. *Substance and Function*, Chicago: The Open Court Publishing Company, 1923.

Coyne, J.A. (1992). “Genetics and Speciation”, *Nature* 355: 511-515.

Collingwood, R. G. (1961). *The Idea of History*. Oxford: Oxford University Press.

Croce, B. (1917). *Logica come Scienza del Concetto Puro*. Bari: G. Laterza.

Danto, A. C. (1965). *Analytical Philosophy of History*. London: Cambridge University Press.

Darwin, C. (1859). *The Origin of Species*. Reprinted 1979, New York: Gramercy Books.

Davidson, D. (1963). “Actions, Reasons and Causes”, *The Journal of Philosophy* 60, 23: 685-700.

Diamond, J.M. (1997). *Guns, Germs, and Steel; The Fates of Human Societies*. New York: W.W. Norton.

Dilthey, W. (1883). *Einleitung in die Geisteswissenschaften: Versuch einer Grundlegung für das Studien der Gesellschaft und der Geschichte; erster band,* Leipzig-Berlin: Duncker & Humblot. Engl. Edit. *Selected Works. Volume 1: Introduction to the Human* Sciences (R.A. Makkreel & F. Rodi, eds.). Princeton: Princeton University Press, 1989.

Dorato, M. (1997). *Futuro Aperto e Libertà. Un’Introduzione alla Filosofia del Tempo*. Roma-Bari: Laterza.

Dray, W.H. (1957). *Laws and Explanation in History*. Oxford: Oxford University Press.

Dray, W.H. (1967). “Singular Hypotheticals and Historical Explanation”, in *Sociological Theories: Inquires and Paradigms* (L.G. Gross, ed.). New York: Harper and Row (181-203).

Eldredge, N., and Cracraft, J. (1980). *Phylogenetic Patterns and the Evolutionary Process*. New York: Columbia University Press.

Farris, J. S. (1973). “A Probability Model for Inferring Evolutionary Trees”, *Systematic Zoology*, 22: 250-256.

Farris, J. S. (1983). “The Logical Basis of Phylogenetic Analysis”. In *Advances in Cladistics,* vol. 2. (N. Platnick & V. Funk, eds). New York: Columbia University Press (7-36). Reprinted in Sober (1984b).

Felsenstein, J. (1973). “Maximum Likelihood and Minimum-Steps Methods for Estimating Evolutionary Trees from Data on Discrete Characters”, *Systematic Zoology*, 22: 240-249.

Felsenstein, J. (1978). “Cases in Which Parsimony or Compatibility Methods Will be Positively Misleading”, *Systematic Zoology* 27: 401-410.

Feyerabend, P.K. (1981). *Philosophical Papers*. Cambridge: Cambridge University Press.

Fitzhugh, K. (2006). “The ‘Requirement of Total Evidence’ and its Role in Phylogenetic Systematics”, *Biology and Philosophy* 21: 309-351.

Forster, M. R. (1988). “Sober’s Principle of Common Cause and the Problem of Incomplete Hypotheses”, *Philosophy of Science* 55: 538-559.

Forster, M. R. (2000). “Hard Problems in the Philosophy of Science – Idealisation and Commensurability”. In *After Popper, Kuhn and Feyerabend* (R. Nola & H. Sankey, eds.). London: Kluwer (231-250).

Forster, M. R. (forthcoming). “A Philosopher’s Guide to Empirical Success”, *Philosophy of Science*.

Forster, M. R. and Sober, E. (1994). “How to Tell When Simpler, More Unified, or Less *Ad Hoc* Theories Will Provide More Accurate Predictions”, *British Journal for the Philosophy of Science* 45: 1-35.

Gaffney, E. (1979). “An Introduction to the Logic of Phylogeny Reconstruction”. In *Phylogenetic Analysis and Paleontology* (J. Cracraft & N. Eldredge, eds.). New York: Columbia University Press (79-111).

Gaddis, J. L. (2002). *The Landscape of History; How Historians Map the Past*. New York: Oxford University Press.

Gannett, L. (2004). "The Biological Reification of Race", *British Journal for the Philosophy of Science* 55: 323-345.

Gardiner, P. (1961). *The Nature of Historical Explanation.* Oxford: Oxford University Press.

Goldberg, D. (1993). *Racist Culture*. Cambridge: Blackwell.

Good, I.J. (1983). *Good Thinking: The Foundation of Probability and Its Application*. Minneapolis: University of Minneapolis Press.

Goodman, N. (1954). *Facts, Fiction and Forecast*, Cambridge (Mass.): Harvard University Press.

Griffiths, P. (2005). "Review of *Niche Construction: The Neglected Process in Evolution* by F. John Odling-Smee, Kevin N. Laland, Marcus W. Feldman", *Biology and Philosophy* 20, 10: 11-20.

Hacking, I. (1965). *Logic of Statistical Inference*. London: Cambridge University Press.

Hacking, I. (1975). *The Emergence of Probability*. London: Cambridge University Press.

Hacking, I. (1981). "Do We See through a Microscope?", *Pacific Philosophical Quarterly* 62: 305-322.

Hacking, I. (2006). "Genetics, Biosocial Groups and the Future of Identity", *Deadelus* 51: 81-95.

Hempel, C.G. (1942). "The Function of General Laws in History", *Journal of Philosophy* 29: 35-48; then in *Theories of History* (P. Gardiner, ed.). Glencoe (Illinois): The Free Press, 1959 (344-356).

Hempel, C.G. (1953). *Reasons and Covering Laws in Historical Explanation*. New York: New York University Press.

Hempel, C.G. (1962). “Explanation in Science and in History”. In *Frontiers in Science and Philosophy* (R.G. Colodny, ed.). Pittsburgh: The University of Pittsburgh Press (9-33).

Hempel, C.G. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press.

Hempel, C.G. and Oppenheim P. (1948). “Studies in the Logic of Explanation”, *Philosophy of Science* 15: 135-175. Reprinted in Hempel (1965).

Hennig, W. (1966). *Phylogenetic Systematics*. Urbana: University of Illinois Press.

Hobsbawm, E. and Ranger, T. (Eds) (1983). *The Invention of Tradition,* Cambridge: Cambridge University Press.

Howson, C. and Urbach, P. (2006). *Scientific Reasoning. The Bayesian Approach*. Chicago and La Salle: Open Court.

Joynt, C.B. and Rescher, N. (1961). “The Problem of Uniqueness in History”, *History and Theory* 1, 2: 150-162.

Kant, I. (1781, 2nd edn 1787). *Kritik der Reinen Vernunft*. Riga: Johann Friedrich Hartknock. Engl. trans. *Immanuel Kant’s Critique of Pure Reason*, London: Macmillan Education, 1989.

Kolmogorov, A.N. (1950). *Foundations of the Theory of Probability*. New York: Chelsea Publishing Company.

Kosso, P. (2001). *Knowing the Past: Philosophical Issues of History and Archaeology*. Amherst (NY): Humanitiy Books.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lakatos, I. (1970). “Falsification and the Methodology of Scientific Reasearch Programmes”. In *Criticism and the Growth of Knowledge* (I. Lakatos & A. Musgrave, eds.). Cambridge: Cambridge University Press (91-196).

Laplace, P. (1814). *Essai Philosophique sur les Probabilités*. Paris: Chancelier. Eng. trans. *A Philosophical Essay on Probabilities*. New York: Dover, 1951.

Laudan, L. (1977). *Progress and Its Problems, Toward a Theory of Scientific Growth*. Berkeley, Los Angeles, London: University of California Press.

Lewontin, R.C. (1982). "Organism and Environment". In *Learning, Development and Culture* (H.C. Plotkin, ed.). New York: Wiley (151-170).

Lewontin, R.C. (1983). "Gene, Organism and Environment". In *Evolution from Molecules to Men* (D.S. Bendall, ed.). Cambridge: Cambridge University Press (273-286).

Lowenthal, D. (1985). *The Past Is a Foreign Country*. Cambridge: Cambridge University Press.

Löwith, K. (1949). *Meaning in History. The Theological Implications of the Philosophy of History.* Chicago: The University of Chicago Press.

Lorusso, L. and Boniolo, G. (forthcoming). "Clustering Humans. On Biological Boundaries", *Studies in History and Philosophy of Biological and Biomedical Sciences*.

Marc-Wogau, K. (1962). "On Historical Explanation", *Theoria* 28: 213-233.

Maddison, W. (1995). "Phylogenetic Histories Within and Among Species". In *Experimental and Molecular Approaches to Plant Biosystematics* (P.C. Hock & A.G. Stephenson, eds). *Monographs in Systematic Botany from the Missouri Botanical Garden*, 53: 289-303.

Mayr, E. (1942). *Systematics and the Origin of Species*. New York: Columbia University Press.

Mayr, E. (1963). *Animal Species and Evolution*. Cambridge (MA): Harvard University Press.

Mayr, E. (1969). *Principles of Systematic Zoology*. New York: McGraw-Hill.

Maxwell, G. (1962). "The Ontological Status of Theoretical Entities" in *Minnesota Studies in the Philosophy of Science* (H. Feigl & G. Maxwell, eds). Minneapolis:

University of Minnesota Press.

Mitchell, S.D. (2002). “Integrative Pluralism”, *Biology and Philosophy* 17: 55-70.

Monod, J. (1970). *Le Hasard et la Nécessité; Essai sur la Philosophie Naturelle de la Biologie Moderne*. Paris: Éditions du Seuil. Engl. Trans. *Chance and Necessity. An Essay on the Natural Philosophy of Modern Biology*. New York: Knopf, 1971.

Nagel, E. (1961). *The Structure of Science*. New York: Harcourt, Brace and World.

Nelson, G. and Platnick, N. (1981). *Systematic and Biogeography: Cladistics and Vicariance.* New York: Columbia University Press.

Neurath, O. (1934). “Radikaler Physikalismum und ‘Wirkliche Welt’”, *Erkenntnis*: 346-362.

Oakeshott, M. (1933). *Experience and Its Modes*. Cambridge: Cambridge University Press. Reprinted 1995.

Orr, H.A. (2001). “The Genetics of Species Differences”, *Trends in Ecological Evolution* 16: 325-413.

Paterson, H.E.H. (1985). “The Recognition Concept of Species”. In *Species and Speciation* (E. S. Vrba, ed.). Pretoria: Transvaal Museum (21-29).

Paterson, H.E.H. (1993). *Evolution and the Recognition Concept of Species. Collected Writings of H.E.H. Paterson* (S. F. McEvey, ed.). Baltimore (MD) : Johns Hopkins University Press.

Pigliucci, M. and Kaplan, J. (2003). “On the Concept of Biological Race and Its Applicability to Humans”. *Philosophy of Science* 70: 1161-1172.

Pigliucci, M. and Kaplan, J. (2006). *Making Sense of Evolution. The Conceptual Foundations of Evolutionary Biology.* Chicago: The University of Chicago Press.

Poincaré, H. (1902). *La Science et l’Hypothèse*. Paris: Flammarion. Eng. trans. *Science and Hypothesis*. New York: Dover, 1952.

Popper, K.R. (1934). *Logik der Forschung*, Wien: Springer. Eng. ed. *The Logic of Scientific Discovery*. London: Hutchinson, 1959.

Popper, K.R. (1944-45). “The Poverty of Historicism”, *Economica* 11, 12. As a book London: Routledge and Kegan Paul, and Boston: The Beacon Press, 1957. Reprinted New York: Harper & Row, 1964.

Popper, K.R. (1945). *The Open Society and Its Enemies*. London: Routledge. Reprinted Princeton (New Jersey): Princeton University Press, 1963.

Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press.

Ridley, M. (1985). *The Problems of Evolution*. Oxford: Oxford University Press.

Ridley, M. (ed.) (2004). *Evolution*. Oxford: Oxford University Press.

Royall, R. (1997). *Statistical Evidence: A Likelihood Paradigm*. London: Chapman & Hall.

Salmon, W. (1975). “Theoretical Explanation”. In *Explanation* (S. Korner, ed.). Oxford: Blackwell (118-145).

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Scriven, M. (1966). “Causes, Connections and Conditions in History”. In *Philosophical Analysis and History* (W.H. Dray, ed.). New York: Harper and Row (238-264).

Shaw, K.L. (2001). “Commentary: The Genealogical View of Speciation”. *Journal of Evolutionary Biology* 14: 880-882.

Simpson, G.G. (1961). *Principles of Animal Taxonomy*. New York: Columbia University Press.

Sneath, P. and Sokal, R. (1973). *Numerical Taxonomy*. San Francisco: Freeman.

Sober, E. (1975). *Simplicity*. Oxford: Clarendon Press.

Sober, E. (1980). “Evolution, Population Thinking, and Essentialism”. *Philosophy of*

Science 47: 350-383.

Sober, E. (1984a). “Common Cause Explanation”. *Philosophy of Science* 51: 212-241.

Sober, E. (ed.) (1984b). *Conceptual Issues in Evolutionary Biology*. Cambridge (Ma): MIT Press.

Sober, E. (1987). “Parsimony, Likelihood, and the Principle of the Common Cause”. *Philosophy of Science* 54: 465-469

Sober, E. (1988). *Reconstructing the Past: Parsimony, Evolution, and Inference*. Cambridge (MA): MIT Press.

Sober, E. (1990). “Let's Razor Ockham’s Razor.” In *Explanation and Its Limits* (D. Knowles, ed.). Cambridge: Cambridge University Press (73-94). Reprinted in E. Sober, *From a Biological Point of View*. Cambridge: Cambridge University Press, 1994.

Sober, E. (2000). *Philosophy of Biology*. Boulder (CO): Westview Press.

Sober, E. (2002a). “Bayesianism – Its Scope and Limits”. In *Bayes' Theorem* (R. Swinburne, ed.). *Proceedings of the British Academy Press* 113: 21-38

Sober, E. (2002b). “Instrumentalism, Parsimony, and the Akaike Framework”, *Philosophy of Science* 69: S112-S123.

Sober, E. (2002c). "What is the Problem of Simplicity?" In *Simplicity, Inference, and Modelling* (A. Zellner, H. Keuzenkamp & M. McAleer, eds.). Cambridge: Cambridge University Press (13-32).

Sober, E. (2004). “The Contest Between Likelihood and Parsimony”, *Systematic Biology* 53: 644-653.

Sober, E. (2005). “Parsimony and Its Presuppositions”. In *Parsimony, Phylogeny, and Genomics* (V. Albert, ed.). Oxford: Oxford University Press.

Sober, E. (2006). “Parsimony”. In *The Philosophy of Science - An Encyclopedia, Vol. 2* (S. Sarkar & J. Pfeifer, eds.). New York: Routledge.

Sober, E. (forthcoming). *Evidence and Evolution. The Logic Behind the Science*. Cambridge: Cambridge University Press.

Stamos, D.N. (2005). "Pre-Darwinian Taxonomy and Essentialism – A Reply to Mary Winsor", *Biology and Philosophy* 20, 1: 79-96.

Stanford, M. (1986). *The Nature of Historical Knowledge*. Oxford and Cambridge: Blackwell.

Timpanaro, S. (1963). *La Genesi del Metodo del Lachmann*. Firenze: Le Monnier.

Tucker, A. (2004). *Our Knowledge of the Past: A Philosophy of Historiography*. Cambridge: Cambridge University Press.

Van Fraassen, B.C. (1980). *The Scientific Image*. Oxford: Oxford University Press.

Weber, M. (1922). *Gesammelte Aufsätze zur Wissenschaftslehere*, Tübingen: Mohr. Engl. trans. *The Methodology of the Social Sciences*, New York : The Free Press, 1949.

Wimsatt, W. (1994). "Lewontin's Evidence (That There Isn't Any)“. In *Questions of Evidence: Proof, Practice, and Persuasion Across the Disciplines* (J. Chandler, A.I. Davidson & H. Harootunian, eds.). Chicago, University of Chicago Press (492-503).

Windelband, W. (1883). *Präludien*. Tübingen: Mohr

Winsor, M.P. (2003). "Non-Essentialist Methods in Pre-Darwinian Taxonomy", *Biology and Philosophy* 18, 3: 387-400.

Wu, C.-I. (2001). "The Genic View of the Process of Speciation", *Journal of Evolutionary Biology* 14: 851-865.