Scientific cognition and cultural evolution: theoretical tools for integrating cognitive and social studies of science

Christophe Heintz

Christophe HEINTZ

COGNITION SCIENTIFIQUE ET ÉVOLUTION CULTURELLE

Outils théoriques pour incorporer les études cognitives aux études sociales de la science

Doctorat nouveau régime
Discipline : PHILOSOPHIE ET SCIENCES SOCIALES

Directeur de thèse :

M. Dan SPERBER, Directeur de recherche, CNRS

Jury de thèse :

M. Daniel ANDLER, Professeur, Université de Paris-Sorbonne

M. Bernard CONEIN, Professeur, Nice-Sophia Antipolis

M. Jérôme DOKIC, Directeur d'études, EHESS

M. David GOODING, Professeur, University of Bath

ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES

Mai 2007

Résumé

Cette thèse préconise d'utiliser des outils théoriques de l'anthropologie cognitive pour l'étude scientifique de la science. Ces outils sont l'épidémiologie des représentations, développée par Dan Sperber, et l'étude de la cognition distribuée, telle qu'elle à été développée par Ed Hutchins. Ces deux théories, qui sont par ailleurs étroitement liées, ont pour apport essentiel de permettre d'intégrer les études cognitives et sociales de la science. Deux études d'histoire des mathématiques illustrent le potentiel explicatif de ces théories : le développement du calcul infinitésimal en France au début du 18ème siècle, et l'avènement des ordinateurs dans la pratique des mathématiques, marqué par la preuve du théorème des quatre couleurs. Les études cognitives de la science commencent depuis peu à prendre au sérieux le caractère social de la production scientifique. Je m'inscris dans les développements récents de ces travaux, tout en soulignant l'apport distinct d'une anthropologie cognitive de la science : considérer la science comme un phénomène culturel, qui par là même implique des processus mentaux, et peut donc bénéficier d'une approche cognitive. La première partie fait un recensement des principes méthodologiques, théoriques et philosophiques qui permettent et trop souvent entravent l'étude intégrée des phénomènes mentaux et sociaux de la production scientifique. C'est finalement les travaux du Programme Fort qui, tout en montrant la nature sociale de la science, se révèle le plus à même d'incorporer des analyses cognitives basées sur des résultats de la psychologie cognitive et évolutionniste. La deuxième partie s'attache aux bases psychologiques de la rationalité scientifique, que les sociologues des sciences ont décrite comme changeante et relative au contexte social. Des théories récentes de la psy-

chologie cognitive montrent que l'esprit humain est richement doté de capacités cognitives universelles fortement déterminées par les contraintes génétiques sur le développement du cerveau. Comment réconcilier l'universalisme de ces capacités psychologiques, qui constituent ensemble la rationalité humaine, et l'observation de la variabilité des normes de rationalité scientifique ? Considérer ces deux rationalités comme distinctes, c'est reconnaître que la nature de l'appareil cognitif humain joue un rôle fondamental dans la formation des croyances scientifiques, mais sous-détermine le contenu de ces croyances. J'analyse les mécanismes mentaux qui permettent aux scientifiques d'inscrire leurs réflexions dans un contexte culturel (par le biais de cadres interprétatifs), tout en exploitant les capacités d'inférences de l'appareil cognitif humain. Ces mécanismes sont présentés comme des moments essentiels dans les chaînes de relations causales de la production des représentations scientifiques. Ils sont situés à l'intérieur de ces chaînes qui comprennent des actes de communication, de perception, et d'expérimentation, et mettent en jeux - et c'est le thème de la partie suivante - de multiples acteurs et artéfacts. La cognition scientifique est distribuée de manière systématique entre des scientifiques spécialisés et des artefacts ; elle est organisée dans les systèmes de cognition distribuée que forment les institutions scientifiques. La troisième partie de la thèse analyse les pratiques scientifiques comme ayant une fonction cognitive au sein de systèmes de cognition distribuée. Comprendre l'évolution de ces pratiques, c'est alors comprendre pourquoi et comment les système de cognition distribuée changent ou perdurent. Je souligne le rôle de certaines représentations qui régulent le fonctionnement des institutions scientifiques en attribuant des fonctions cognitives à certains éléments, qu'ils soient humains ou non. Ces 'représentations régulatives' sont en fait des croyances sur les moyens de production de la connaissance scientifiques : l'organisation des institutions scientifiques évolue lorsqu'un scientifique se propose d'utiliser les représentations produites par un élément cognitif qu'il juge suffisamment fiable. Une étude épidémiologique sur la distribution des représentations de fiabilité cognitive contribue à expliquer l'évolution des institutions scientifiques, et par ce biais, l'évolution des pratiques et de la connaissance scientifique. Il y a donc différents mécanismes qui distribuent les représentations scientifiques, pour générer le phénomène culturel qu'est l'évolution de la science. Je m'attache dans cette thèse à décrire certains mécanismes mentaux dans la formation de croyance des scientifiques et certains mécanismes sociaux dans la formation des institutions scientifiques. Mais il se révèle que ces mécanismes s'inscrivent toujours à l'intérieur de chaînes causales qui incluent processus mentaux et interactions sociales. Des études intégrant les aspects cognitifs et sociaux sont donc nécessaires ; et je tente de contribuer à leur développement en articulant théories de sociologie des sciences et théories de psychologie cognitive dans le but de mieux appréhender l'histoire des sciences.

Table des matières

Christophe HEINTZ

Scientific cognition and cultural evolution:
theoretical tools for integrating social and cognitive studies of science

Ph.D Thesis, EHESS, Paris
May 2007

Keywords: cognition and culture, science—social aspects—cognitive aspects, historiography, history of mathematics—case studies

An electronic version of this document is available at:
http://christophe.heintz.free.fr/thesis

Contents

Preface

I began to be interested in the topic of rationality when following the courses of Raymon Boudon in sociology at the Sorbonne. At the same time, I was attending the courses of Alban Bouvier in the sociology of science, and it seemed a good idea to investigate the problem of rationality through the study of scientific thinking and practices. At Cambridge, Martin Kusch made me discover the sociology of the Strong Programme, and, although I arrived with strong rationalist presuppositions, I was convinced that the Strong Programme was advocating a proper method for the naturalistic study of science. As I manifested my interests in the cognitive aspects of scientific knowledge production, Martin Kusch advised me to read Sperber's *Explaining Culture* and Atran's *Cognitive Foundations of Natural History*. With the Strong Programme and the Epidemiology of Representations, Martin Kusch gave me the ingredients for my own work. I thank him very much for that. In many aspects, the present thesis is a straightforward application, to the history of science, of Dan Sperber's ideas about the cognitive foundations of cultural variability and stability—this is, at least, how I think of it. The added value I bring in may be due to my interest in the 'relativist' sociology of scientific knowledge.

I thank very much Dan Sperber, my supervisor. His targeted comments and more general discussions have been of great help to me.

Many thanks to the people who have read and commented parts of this thesis. They are: Nicolas Baumard, Valeria Giardino, Paul Egre, Hugo Mercier, Olivier Morin, Nancy Nersessian, Dario Taraborelli. Davide Vecchi, I also benefited from discussions at the Max Planck Institute for the History of Science, where I spent six months as a visiting fellow, and at the Institut Jean Nicod, my home Institute.

I thank my parents and my wife, Monica Heintz, for their help and support. Monica took on herself to support financially our children and her husband, and she did all she could to give me the best conditions of work.

Part I

Motives and means for integrating cognitive and social studies of science

Chapter 1

Introduction

The evolution of science is an instance of cultural evolution. Science is a historical product of interacting people; and it is a cultural phenomenon constituted of an evolving distribution of representations among the scientific community. The evolution of science is also a cognitive process, as it involves the production, transformation and distribution of representations. These representations include, but are not restricted to, scientist's beliefs; scientists' mental processes are prominent causes of scientific evolution.

The present work is a study in the historiography of science. Its goal is to provide some theoretical tools for studying the evolution of science as a social and cognitive phenomenon. It aims at showing that some concepts and frames of analysis drawn from cognitive anthropology are fruitful tools for the scientific study of science. The theories that I advocate using are *the epidemiology of representation* and *the theory of distributed cognition*. The added value of these theories stems in great part from their enabling to integrate results from cognitive and social studies of science.

Science studies is an interdisciplinary field, which include subfields in history, sociology and anthropology, psychology and cognitive science, and philosophy, and which questions what should be the relation between its disciplinary constituents. For instance, there are numerous departments of 'History and Philosophy of Science,' but the relation between normative epistemology and empirical studies of science is much debated.

This work, however, is concerned with the relations between the empirical sciences of sciences. The main idea is that integrating cognitive and social studies of science would importantly advance the naturalistic programme in epistemology, which aims at developing a scientific account of scientific development. Its goal is therefore to articulate approaches and results from cognitive psychology and sociology, in order to describe the causal relations between the mental and social events that make scientific knowledge. The first step for achieving this goal is to de-dramatise the opposition and tensions that exist between social and cognitive studies of science; the second step is to find some theoretical means for calling on theories that come from both psychology and sociology in the explanation of scientific historical events; the third step is to select the best social and psychological theories, work out their compatibility ... and put them at work in case studies.

1.1 PUTTING THE SOCIAL AND THE COGNITIVE TOGETHER AGAIN

The project of integrating social and cognitive studies of science is not self evident, as these two studies sometimes present themselves as incompatible with one another. One basic assertion of sociologists of science is that it is a social phenomenon. This credo faces resistance from the renewed assertion of the special status of science as a rational enterprise (e.g. Laudan 1990, some protagonists of the Science War). From this, some have concluded that only cognitive studies can reveal the real nature of science, that scientific discovery by computer was a refutation of the Strong Programme in the sociology of scientific knowledge (Slezak, 1989), and some have emphasised that “cognitive theories do not agree with the relativist epistemology advocated within the sociology of knowledge” (Freedman, 1997). Often, the attitude is simply dismissive. In their ‘very short recent history’ of the philosophy of science, Carruthers et al. (2002) dedicate a section to ‘Science and the Social,’ which they introduce by:

> our story so far has mostly been of good news – with philosophy of science in the last century, like science itself, arguably

> progressing and/or getting somewhat closer to the truth. But one outgrowth of the historical turn in philosophy of science was a form of social constructivism or relativism about science (Bloor, 1976; Rorty, 1979; Latour and Woolgar, 1986; Shapin, 1994) [. . .] social constructivism has not found wide acceptance among philosophers of science generally (nor among the contributors of this volume in particular).

While these reactions have mostly targeted social studies of science, cognitive studies of science have not been spared from doubts and criticism. From the social studies of science, the doubts and criticisms have been most dramatically represented by the notorious call for a moratorium on cognitive studies of science (Latour & Woolgar, 1986: 280; Latour, 1987: 247). The positive result of these criticisms is the specification of two extreme opposite positions to avoid: a sociological view according to which scientists' minds do nothing but reflect or express the social context; and a cognitive view that would transpose all the normative laws of the positivists into the scientists' minds. The relativism *versus* rationalism debate has been the main cause of disagreement between sociologists, advocates of relativism, and the proponents of a cognitive approach, defenders of rationalism. One strategy for enabling integrative studies of science is to reconcile the rationalist position and the relativist sociology of science through compromises: the rationalists may not be entirely true and that the relativists may have gone a little too far. Integration, however, is not desired for the sake of syncretism, but because systematically ignoring the results of either psychology or sociology in science studies goes against the naturalistic programme in epistemology: developing a science of science.

1.2 Means of integration

Theories specifying the relations between the cultural and the mental phenomena enable articulating social and psychological theories of science. The theories I draw upon come from cognitive anthropology. Science is then considered as a cultural phenomenon which, as any cultural phenomenon, implicates mental processes. The main idea of cognitive approaches in anthropology is that cognitive science should be taken seriously. This implies that explanations of socio-cultural phenomena should not be blatantly contradicting findings in cognitive science, and it suggests that some explanations may include factors that have been specified by cognitive science. There are different trends within cognitive anthropology, depending on the cognitive theories drawn upon and how culture is defined. One classical view uses the theories of mental models, schema and scripts, and takes culture as being constituted by the knowledge people must have in order to manage in a community. Another view is represented by Sperber, Boyer, Atran, Hirschfeld and others, who analyse cultural phenomena as relatively stable distributions of representations, mental and public, among a population. They argue that the mind is endowed with innate cognitive abilities which contribute to explaining both cultural stability and diversity. It is this latter approach that I will be advocating as applicable in science studies. The history of science is then seen as an evolution in the distribution of representations within the scientific community—with new scientific representations replacing old ones; and human cognitive capacities are given their own causal roles in the distribution process.

An important challenge in naturalistic epistemology is the apparent immaterial nature of knowledge, which seems to subtract knowledge from causal accounts of it. The cognitive revolution was driven by the will to specify how cognition could be materially implemented. The ensuing research programme first focused on Artificial Intelligence research, but involves more and more interdisciplinary work between brain sciences and psychology. Works in embodied, situated and distributed cognition further add that cognition can be implemented in actions, in cognitive tools and across people. Social interactions, in particular, involve flows of information and transformation of representations. Sperber and Hutchins' theoretical framework enable analysing cognition as it is implemented in social interactions. Hutchins' method of analysis, the analysis of distributed cognition, includes specifying what the cognitive elements are, what are are their cognitive functions, and how they are organised to produce the required cognitive output. This analysis is applicable in science studies, and brings forward the explanatory power of cognitive science in the analysis of social phenomena. The analysis of distributed cognition also provides a way to specify the place of mental processes among the cognitive processes of the encompassing social institutions that produce science.

1.3 Theoretical claims about the principles of scientific evolution

Choosing the theoretical resources for studying science requires selecting among important and good candidates, as concurrent theories abound in psychology and sociology. Some social theories explicitly block psychological inquiries, or they remain at the macro-social level, which makes the cognitive agent disappear; yet some other theories rely on false psychological assumptions. I will advocate the methodology of the Strong Programme in the sociology of science for its integrative potential, and I will mostly draw on its theories. As for theories in psychology, I will use computational psychology, the theories of domain specific competencies, and evolutionary psychology. One reason for choosing these psychological theories for science studies is that they have been elaborated with motivations independent from the fact that some few people do science. I argue that they can nonetheless account for scientific reasoning, which strongly corroborates them. The main motive for my choices of theories is their naturalistic potential, as they seek to develop causal accounts of human behaviour, relating as much as possible their theoretical claims to the natural sciences, from the analysis of social phenomena, to the study of the cognitive processes causing individuals' behaviour, and to the evolved biological bases of the human cognitive apparatus. To the idea of a disembodied rationality coming from nowhere and compelling scientific thinking through no clear means, the naturalist substitutes the analysis of brain processes and socially implemented norms of reasoning.

Although my main point is methodological, I argue in favour of integration by actually putting some theories together. In doing so, I develop an account of the principles of scientific evolution. Scientific thinking relies on the cognitive capacities with which humans are endowed, and it is also socially situated. Scientific thinking is reflexive, which means that intuitive representations are reflected upon—they are meta-represented. In the process, scientists take their own context into account: it includes the current accepted theories, the existing means for convincing others, and it may include many other factors. This contextual knowledge forms the interpretive framework with which new information is processed in the formation of scientific beliefs. Metarepresentational abilities provide the cognitive foundations of the Strong Programme's analysis of the contextual aspects of scientific belief formation. However, mental competencies other than metarepresentational have their own roles in scientific thinking, as they provide the inferential power and intuitive beliefs with which scientific representations are processed. Both contextual factors and the innate human cognitive endowment play a role in the evolution of scientific beliefs. In particular, representations that make better use of human cognitive abilities, current knowledge and the cognitive environment, produce more information at a lower cost. They are consequently more 'attractive' than other concurrent representations, and have more chance to stabilise and acquire the status of scientific representations. Meta-representations trigger cognitive abilities by meeting their input conditions, thus enabling their cultural cognitive exploitation. The tendency to make the most of cognitive elements is generalised to the exploitation of cognitive artefacts and the organisation of cognition in social institutions. Cognition is also implemented in the external environment and in scientists' actions. The analysis of distributed cognition in science reveals that the actual producers of scientific knowledge are well designed distributed cognitive systems. The causes of the specific organisation of these systems are historical. Distributed cognitive systems evolve through changes in the distribution of representations that attribute cognitive functions to elements: when cognitive elements are trusted for some cognitive task, they are given a place within some system. There are numerous mechanisms that participate to the production and distribution of scientific representations. These mechanisms necessarily involve human innate cognitive abilities, but as science involves complex cognition and hard to understand representations, it also importantly involves situated and distributed cognition, organised in social institutions. The construction of the scientific environment is a key means of scientific evolution, as "humans create their cognitive power by creating the environments in which they exercise those powers" (Hutchins, 1995, p. 169). One of my points is that Hutchins claim well characterise the specificity of scientific cognition, which is empowered by a constructed cultural, social, and material environment.

1.4 Organisation of the thesis

The thesis is organised in four parts. The first part exposes the motives in favour of an integrated causal model of scientific evolution and reviews the available means. In the next chapter, I critically review some integrated models of scientific evolution. Then, in chapter three, I argue that the Strong Programme is the best starting point for integrated, causal, studies of science.

Parts two and three expose respectively the epidemiology of representations and the theory of distributed cognition and show how they apply to the analysis of science. Each part includes three theoretical chapters and one chapter illustrating the approach with a case study in the history of mathematics.

Chapter four, the first chapter of part two, introduces the problem of rationality as it occurs in science studies and proposes some solutions by distinguishing epistemological beliefs and knowledge practices on the one hand, and human cognitive abilities on the other. It proposes to study how the latter enables and constrains the formers. Chapter five expands on the innate endowment of the human mind and its consequence on the evolution of science. It shows how to accommodate nativism and some type of social constructivism in science studies. With these theories in hand, I present, in chapter six, the epidemiology of scientific representations, leading me to question the specificity of some mechanisms of the production and distribution of scientific knowledge. I apply the theoretical framework to a historical case study: the epidemiology of the 'infinitesimals' representations — especially as the occurred at the beginning of 18^{th} century France. Through the epidemiological approach, I can draw hypotheses on the cognitive bases of the calculus and its causal role on the development of mathematics without being 'psychologistic.'

The third part is devoted to the theory of distributed cognition as applied to scientific practices. I first expose the theory and review works in science studies which applied it. The question I especially deal with is the evolution of distributed cognitive systems: chapters eight and nine attempt to find the principles through which distributed cognitive systems in science are organised, maintained through time, and changed. I eventually argue that the distribution of beliefs about what is trustworthy, and for which task, is critical in the evolution of distributed cognitive systems. Chapter eleven illustrates this claim with the analysis of the advent of computer assisted proof, with the four-colour theorem.

The fourth part is made of a conclusive chapter: I contrast the mechanisms of scientific knowledge production and distribution that I have specified in the thesis with attempts of evolutionary epistemology to reduce such mechanisms to blind variation and selective retention. I also point out some consequences of socio-cognitive studies of science for cognitive psychology.

Chapter 2

Cognitive studies of science and the social

When Quine (1969), in his seminal *Naturalizing Epistemology*, explained that epistemology should make proper use of the natural sciences in its attempt to explain what science is, he directed towards psychology. He more radically asserted that "epistemology, or something like it, simply falls into place as a chapter of psychology." Yet, the naturalistic programme—explaining science as a natural phenomenon with the means of science itself—was mostly took on by *social* studies of science, which issued a literature much larger than the one in the psychology of science. The naturalisation of epistemology happened through two different approaches: one emphasising the contingent social factors in scientific knowledge making, the other attempting to grasp the natural—psychological—basis of scientific thinking and reasoning. Even though the two approaches have had the same naturalistic goal, they remained largely distinct. There are some good reasons for that: each approach has had its specific object and its own methodology. Initially, psychology has experimentation as its main tool: subjects come to a laboratory where they pass some test that psychologists analyse. Cognitive studies investigate phenomena that are located in the mind/brain, such as individual's processing of information and their cognitive faculties; it assumes that there are universal features of human cognition. By contrast, sociology mainly uses statistical data such as gender rates, historical records and textitin situ observations. The sociology of knowledge investigates the social causes in the making of knowledge; it focuses on the collective construction of knowledge and is keen to point out the diversity of cultural representations. While cognitive psychology mainly talks of computational processes, the sociology of knowledge frames its explanation in terms of social choices and strategies. In spite of these genuine and comprehensible differences, there are very good reasons why we should not remain with such a simplistic compartmentalised view. In actual research, the difference between the cognitive and social approaches is, fortunately, far from being this clear cut.

In this chapter, I advocate bridging cognitive studies to social studies of science (in the next chapter I will do the converse: advocate bridging social studies to cognitive and psychological studies of science). The underlying idea is that science *is* a historical and social phenomenon. Cognitive sciences can better contribute to the understanding of science when envisaged as specifying the mental and cognitive processes at the basis of this socio-cultural phenomenon. This implies, in particular, situating scientific thinking in its cultural context: asking how the context may have contributed to the advent and content of the thoughts, and reciprocally, how the thoughts have contributed to change the cultural scientific context. More generally, this implies integrating psychological and social studies of science.

In the first section of this chapter, I briefly spell out some reasons why integrating cognitive and social studies of science is a worthwhile project. I then denounce some ideas as hindering integrations and reinforcing the unproductive divide between cognitive and social studies of science. Integrative research, however, have mainly stemmed from the psychology of science, rather than from social studies; I inscribe my own project within this new and ongoing research. In the last section, I point out the contribution that cognitive anthropology could have to the integrative project.

2.1 Why integrating social and cognitive studies of science

Psychology for the social sciences

One reason for integrating social and psychological studies of science regards the relation between sociology and psychology in general. The naturalisation of the social sciences is itself a research programme that requires specifying the natural, or material, basis of social phenomena. This implies rethinking social phenomena as arising from people's behaviour, which is determined by their thoughts, which are studied in psychology [1]. Naturalising the social sciences implies understanding the mental causes of the behaviours that cause social and cultural phenomena. It implies bridging psychology and the social sciences. My contention is that the naturalistic programme in epistemology does include the naturalistic programme in the social sciences. I have two assumptions: (1) sociology is an inescapable discipline in the naturalisation of epistemology, a point that is much defended by sociologists of scientific knowledge, and that I will review in the next chapter (chap. 3); (2) psychology is itself essential in the naturalisation of the social sciences: this point is defended all along the thesis, especially in chap. 5 (see also chap. 6 and 10 about the role of mental representations). The naturalistic programme in epistemology eventually aims at finding out how a world, which is as described by the natural sciences, with the causal relation and natural entities they posit, eventually include humans doing science. I will argue that sociology cannot be bypassed in this research enterprise; psychology neither, since one must be able to specify the psychological fabric of social phenomena.

the social determination of psychological processes

Another good reason for making connections between social and psychological studies of science stems from the extensive scope that the sociology

[1] note that this programme is not necessarily 'reductivist' as the term is used in derogatory arguments: the usefulness of an explanatory level that uses social entities is not denied. The attempt is not to replace sociological explanations, but to investigate the natural status of the social entities involved in explanations.

of scientific knowledge claims to have: it argues that what and how scientists think is determined by the social context. Sociologists of scientific knowledge (esp. sociologists of the Edinburgh school) claim to have essential information for understanding the problem of scientific belief formation, even though it is normally thought of as a psychological process. So it is sociology, here, that is made relevant for psychology. How do social phenomena act upon scientific thinking, and to which extend do they determine it? I will argue, for instance, that scientific enculturation has little effect on the architecture of the mind (see section 12.2.2) but that it nonetheless determines scientists' interpretation of observed data (see section 4.2.2). Psychology and sociology meet at numerous points in science studies: with social psychology of science, in the study of scientific enculturation and of the communication of scientific results, and with questions dealing with interpretations in scientific controversies, where both thought processes involved in understanding and social background have a role. Historians of science have known for long that the texts and other representations of scientists need to be contextualised in order to grasp the authors' intentions and ideas — the productions and decisions of scientists are context dependant in some non-trivial ways.

The encompassing concept of cognition

One last reason for connecting social and cognitive studies of science is provided by the encompassing concept of cognition: information is not processed in minds only, but also in the external environment, possibly shaped by human action. In particular, scientific practices make extensive use of tools that, as they transform information, qualify as cognitive devices. Scientific production is mainly cognitive: it consists largely in public representations (e.g. journal articles) that are further processed by scientists. Information flow in and out people's head, through external media, and information flow is constrained by the social structure. Cognition is a concept that can be used to describe both events in the heads and events in society; the concept is appropriate as soon as the events involve representations or information being processed. This motivates larger frames of analysis that encompass events happening in the brain and social events (see section 2.3.2).

Pluralism of methods for the naturalistic study of science

Fortunately, the complexity of the relation between social, cognitive and psychological events is already reflected in the methods of analyses, with methods falling in between sociology and psychology. Science studies – because it is an interdisciplinary field – has numerous methods of investigation at its disposal. Klahr and Simon (1999) distinguish historical studies, laboratory studies, direct observation and computational modelling, and show how they complement each other. Dunbar and Fugelsang (2005) add the study of brain patterns using techniques such as magnetic resonance imaging and use different names: they distinguish in vitro (laboratory studies), ex vivo (direct observation), in silico (computational modelling), in magnetico (study of brain patterns), sub specie historiae (historical studies). The main method of anthropology of science is direct observation or *participant observation*: the observation of scientific practices in real life settings by being present in these settings and participating in the research activities (Latour, 1986). Dunbar (1995) notes that anthropological studies of this kind are not normally concerned with the cognitive processes that are used by scientists in their day-to-day research. He thus introduced the term *in vivo* studies for ethnographic investigations focusing on scientific cognition and showed its fruitfulness, leading to findings complementary to laboratory, in vitro, studies (Dunbar, 1995; Dunbar and Blanchette, 2001, e.g.). For instance, while in vivo studies allow discovering the causal role of the context in the generation of analogy, in vitro studies allow further identification of which aspects of the context have a causal role. In parallel, Hutchins has advocated a method of investigation—cognitive ethnography—that is adapted for the analysis of cognition in the wild, as embodied, culturally immersed and socially distributed. Cognitive ethnography describes what are the cognitive tasks of a system and its elements, focuses on events, and, most importantly, brings together relevant techniques for achieving its goal, including interviews, surveys, participant observation and a special attention to video and audio recording. So the difference between in vivo study and cognitive ethnography, if any, is that the former emphasises the reasoning practices of scientists in research situations while the latter privileges the description of external cognitive processes such as the manipulation and transformation of external representations. Yet another method of investigation is cognitive history of science as defined by Nersessian (1995): joining "historical enquiries with those carried out in the sciences of cognition." (see section 2.3.1)

In spite of all these reasons for integration, there remain two very distinct approaches in the sciences of science, one psychological and cognitive, and the other one social. More often than not, cognitive psychologists accuse sociologists of knowledge of the sin of relativism, and of missing the essential attribute of cognition, namely to grasp what is the natural world. Sociologists of scientific knowledge have been fighting so much to show that knowledge should really be their object of study that those most open to psychological studies have nonetheless completely neglected its possible resources. Often, sociology of scientific knowledge has been threatening, or has been understood as threatening, cognitive and psychological studies, with claims taken to be equivalent to "everything is social." At best, sociology of scientific knowledge has been unclear about the role it 'leaves' to cognitive and psychological studies of science.

2.2 IMPEDIMENTS TO INTEGRATION

The first step for promoting integrated studies of science would be to end the unproductive divide between cognitive and social studies of science. While the cognitive revolution and the historical and social turn in science studies both took place some forty years ago, the two naturalistic trends did not really meet to form an integrated framework. It is only in 1989 that Fuller et al. edited a book on the social and cognitive perspectives in science sciences. The book has the promising title 'The Cognitive Turn: Sociological and Psychological Perspectives on Science', but is presented with the following description: "our historians, philosophers, sociologists, and psychologists [...] have drawn new lines of battle." What a scientific achievement! The book, indeed, is more confronting perspectives than paving a road to fruitful integrated studies.

The opposition between sociological and cognitive approaches to science is in great part undermined by the old rationalist *versus* relativist opposition. Insofar as the debate takes place within a philosophical foundationalist perspective—Is there some rational foundation of science?—it is irrelevant to the issue of integration. For an empirical science of science, the objective is clear: study the causal chains leading to science independently of one's own assessment of the practices or beliefs under investigation. The true question, for an integrative project, bears upon which theoretical resources are appropriate for guiding the empirical studies of science (Downes, 2001, p.227). Answers to these questions can be surprisingly radical: most authors in sociology of science think that cognitive psychology is little or not relevant for the study of science. One big impediment to integration comes from sociologists' ban on psychological studies; I will extensively argue that the ban is unjustified and unfruitful (see chap. **??**). But philosophers or cognitive scientists of science have enounced similar bans on current research in the sociology of science, arguing either that there could not be social determination of the content of scientific knowledge, or dismissing the post-Mertonian sociological research programme (see e.g. quote in section 1.1, Slezak 1989; Bricmont and Sokal 2001; Laudan 1987, arguments against the strong programme as a research programme). I will argue that this ban is ill grounded, because it based on the erroneous thought that the dismissed sociological programmes negatively evaluate science. The programmes, however, attempt to study science independently of epistemic evaluations. Renewed epistemic evaluations, aimed at debunking sociological studies, do not serve the integrative programme.

2.2.1 PHILOSOPHERS' QUALMS ABOUT SOCIAL DETERMINATION

Ronald Giere, Philip Kitcher and Alvin Goldman are philosophers of science who have largely drawn on cognitive science in order to give a natural account of scientific thinking. They have also advocated the integration of social studies and cognitive studies of science. Yet, they always introduce their social studies with strong criticism of the sociology of scientific knowledge of the Strong Programme. I will argue that their criticism is out of the target and stems from their persistent will not to undermine the alleged rational foundations of science. This motivates them to deny, without empirical arguments, that social phenomena importantly enter the processes of scientific belief formation and act on the content of scientific knowledge.

Giere, Kitcher and Goldman support a weak programme in the sociology of scientific knowledge, which consists in tracing and analysing social and political factors acting on the institutions of science. It investigates social causes as determining the career choices of scientists, or the funding of projects, or research agenda. The further assertion of the Strong Programme is that social events have causal power not only on the orientation or developmental speed of scientific research, but also on its very content. Social events determine hypothesis formations and justifications of truth assertion. They are, moreover, constitutive of what transform individuals' beliefs into recognised scientific knowledge. In view of these changes, from weak to strong, Kitcher (2000) claims that "the sociology of science is moribund", that it has "fallen on very hard times" since Robert Merton's weak programme. His argument consists in praising some works in sociology and dismissing Shapin and Schaffer's 1985 book. The book, Kitcher says, does not reveal the sins of methodological individualism (but, actually, it did not intend to) and is just social history (which is what it was meant to be). I think that such strong reactions are based on a misunderstanding of the Strong Programme by the psychologically oriented philosophers of science: The Strong Programme is thought as a 'weak programme' in the sociology of science that would claim that those social and political factors that are unrelated to processes of belief formation are all that there is in the development of science. Philosophers are too often satisfied with mocking and flogging this straw-man. Goldman (1994) begins with the following account of what he takes to be the sociological theory of "the dominant approach to persuasion in the social studies of science": "Scientists are persuaded by the force or weight of greater numbers" (p.278), which becomes, after Goldman charitably enriched it: "What causes a hearer to believe or disbelieve a speaker's claim are the hearer's political or professional interests". The dominant approach is then dismissed with the plain "one cannot believe a desired conclusion simply by an act of will". In favour of Goldman, we can notice that his argument is based on Latour's writing, which I do not intend to defend here. It reveals, however, another repeated flaw in the argumentation: the different paradigms in the social studies of science are often not distinguished by the opponents, who then take for building their straw man, one assertion here, one assertion there. The regrettable result of this not thought enough opposition is that the integrative project of these authors is deprived of most important resources in sociology. Turning their back to sociology, Goldman (1992, 1994, 1999) and Kitcher (1992, 1993) find the resources for integration into economics. The economic model could prove very fruitful and Goldman and Kitcher's integrative project is among the most elaborated attempt. There are nonetheless reasons why their project is misguided: their bias in favour of the rational scientist unstained by the social, lead them to develop an unrealistic account of the scientists' though processes.

2.2.2 Homo-economicus, homo-scientificus

There are, in fact, two trends of work in Goldman and Kitcher's writings. One trend qualifies as naturalised individualistic epistemology (Goldman, 1986, see in particular), which makes great use of the results from cognitive psychology. For instance, Kitcher (1993, p.66) cites Goldman as one of the main philosophical sources for discussion of the inapplicability of

Bayesianism in epistemology. The other trend of work qualifies as social epistemology. In this trend, Goldman and Kitcher replace their source of insights about agent's thinking processes, from cognitive psychology to neo-classical economics. Goldman and Kitcher's economic model take agents to be fully rational, in accord with Rational Choice theory and neo-classical microeconomics. The decisions of Goldman and Kitcher's agents are thus guided by principles of probability and decision theory. For instance, Goldman uses Bayes theorem for modelling the belief revision caused by hearing a testimony. The point against Goldman and Kitcher is that this economic model of the rational agent is not empirically viable. This point seems to have been acknowledged by the same authors when doing in individualistic epistemology.

Cognitive psychologists and behavioural economists, most notably Tversky and Kahneman (eg. Kahneman et al., 1982), have shown that economic agents do not reason or behave in the way predicted by the rational agent model. Pointing out the contradiction in Goldman and Kitcher's work, Downes (2001) notices that the two philosophers have themselves asserted the importance of Tversky-Kahneman results for epistemology. Downes uses the distinction between thin and thick concepts of agents: the former are ideal utility maximizer, while the latter endows agents with a much richer psychology – such as a rich cognitive structure – and conceive them as interpreting their environment. Thick concepts of agents are present, he says (pp. 229-30), in anthropology and cognitive psychology, and consequently in Goldman and Kitcher's individualistic epistemology. Bloor also falls in the category of 'having a thick concept of agent' because he endows scientists with motivation and interpretive abilities informed by the context. This holds although Bloor did not investigate the empirical cognitive foundations of his concept of agent. At the opposite side, Latour, neoclassical economics, and Goldman and Kitcher's social epistemology use only a thin concept of agent. And yet, Kitcher (2000, p.36) accuses Shapin and Schaffer of treating "folk views of human motivations as resources" before turning to an even thinner concept of agent! Why do Kitcher and Goldman choose to use the rational choice paradigm, when economics itself is oriented towards a behavioural economics that aims at modelling the economic agent with findings in cognitive psychology and experiments on subject's economic reasoning? Why are Goldman and Kitcher holding different conceptions of agent when studying the psychological and the social aspects of scientific knowledge production? I will argue that, in fact, using the rich resources of cognitive psychology in social epistemology would inevitably lead to the sociologistic views that they dismiss (see esp. chap. 4). Using a ready made rational agent in order to account for social interactions blocks the analysis of the social constitution of norms of rationality or the context dependence of processes of scientific belief formation.

2.2.3 When foundational concerns get in the way

A better understanding of Goldman and Kitcher's choice is, however, to be found in their project, which can be characterised as naturalised normative epistemology. Kitcher and Goldman aim at prescribing rules or norms for the improvement of science's methods given that (1) the "the aim of inquiry [..] is to obtain *significant* truth" (Kitcher, 1992: 102) and (2) the attainment of this goal is to be done with our human limited abilities. The first point is presented as the justification of the normative project: it is because there is a goal that is essential to science that we can devise norms that are not relative to limited contexts of enquiry. The second point provides the reason why the findings of sciences should be taken into account: one needs to know how scientists actually produce scientific knowledge so as understand what would improve the practices. This requirement has been put to the forefront with the failure of post-fregean epistemology to provide fundamental principles for scientific investigation. Such principles, in fact, need to be specified in the light of empirical investigations and are, consequently, revisable. Goldman and Kitcher are ultimately more preoccupied with normative epistemology than with the natural science of science. Their epistemology is naturalised because it elaborates its norms of rationality on the findings of natural sciences rather than, as with post-fregean epistemology, on *a priori*, foundational, philosophical investigations. Goldman and Kitcher's framework – reliabilism – is oriented towards normative prescriptions. When choosing the theoretical resources for social studies of science, they have picked up the one that served better their normative goals rather than the one that could best describe the social character of scientific practices. It is of course sociology that describes best the social character of scientific practices, but its investigation led to a relativisation of norms of reasoning that do not fit Kitcher and Goldman's prescriptive ambitions [2]. Moreover, the sociology of scientific knowledge of the Strong Programme has taken truth claims as a main object of investigation and has warned that the truth and falsity of the claims to be accounted for could not enter in a naturalistic explanation of why these claims were made, because it would lead to teleological explanation and conceal the actual causes. Kitcher and Goldman's framework, by contrast, has truth as a central primary notion, evaluations and prescriptions are aimed at increasing the production significant truths, and the productivity of scientific practices (see, e.g. Goldman, 1999, part 1); the framework more evaluative than explanatory.

Hence the second reason why their project is misguiding for someone aiming at developing an integrated science of science: their normative framework is ill suited for a radically naturalistic, descriptive and explanatory project. This by itself is not a problem, if the framework is not put at work for an integrated science of science. But as Goldman and Kitcher use empirical sciences for their normative purposes, they also constitute their own description of scientific practices with the same framework destined for their meliorative project. Thus, Kitcher (1992) says: "If epistemic appraisals play a role in understanding the history of science, it is because we hope to defend science as a privileged tradition, one that is more worthy of trust and of social support than rival traditions or institutions" (p. 68, note 44). This sentence well describes what I take to be an initial method-

[2] one consequence of relativism is that normative prescriptions cannot pretend to be universally valid, they may be locally satisfactory only. It implies having a modest stance when doing normative epistemology, but it does not forbid giving prescriptions.

ological mistake in Goldman and Kitcher's analysis. There are different kinds of naturalised epistemology depending on the roles that are given to natural sciences, conceptual analysis, normative and evaluative projects and Kitcher (1992) nicely situates his and Goldman's position. But his argument is that empirical, descriptive projects are to be used for normative ones, while they also do the reverse: taking for granted that science is a worthwhile enterprise (epistemic appraisal), then attempting to show that this is the case. In the name of the objectivity, independence and neutrality of scientific investigation, it is acknowledged that scientific inquiries should be as independent as possible of evaluative concerns. So, "epistemic appraisals" should play no role in understanding the history of science. In other words, scientific norms, including norms of good reasoning, are, in a science of science, what need to be explained rather than figuring in the explanans. Because the meliorative goal is not sufficiently distinguished from the scientific one, Goldman and Kitcher are biased in their descriptive assertions. With this emphasis on the empirical ground of the theory, on the one hand, and the 'epistemic' goal already in sight, on the other, the project is reminiscent of scientific socialism. In the end, Kitcher (1993) provides a picture of scientific practices where Adam Smith sides with Hegel for showing that social interest and selfishness are just means for the advancement of Reason. But his picture of the evolution of science relies on a denial of the causal action of social factors on epistemic thoughts. Epistemic thoughts consist in evaluating evidence or argumentation, thus leading to 'doxatic attitutes' such as believing or accepting a theory. Now, Goldman and Kitcher have been working hard at building a strong rampart around the 'epistemic'; they have been attentive to protect this cognitive heart from the pervasive invasion of social factors. The 'epistemic' is then free to dictate alone what really is Truth and Rationality.

The unjustified and protective separation of the social and the epistemic is similarly present in Giere's (1988) re-introduction of social factors in his cognitive epistemology. Nonetheless, tracing the complex interactions of social and cognitive causes should be the focus of an integrative project. Thus Goldman begins his article 'Psychological, Social, and Epistemic Factors in the Theory of Science' (1994) with a rich and promising definition of the term 'social':

> It applies, first, to any causal interactions between two or more agents. When opinions or behavior of certain scientists influence the opinions or behaviour of others, this qualifies as a social transaction. Second, the term 'social' applies to a single agent's psychological state if the state's propositional content refers to the actions or attitudes (actual or prospective) of other agents. A scientist's belief about the beliefs of other scientists, for example, is a social belief, i.e., a belief with a social content. Similarly, a desire to persuade other scientists that one has made a significant contribution is a social desire. In this usage there is obviously no incompatibility between psychological and social factors, since many states turn out to be both psychological and social. A third use of 'social' pertains to institutional rules, such as codes of professional conduct or guidelines for awarding a prize. A fourth sense of 'social' refers to global properties of a community, e.g., the distribution of beliefs in a population at a given time, whether of not there is causal interaction or intentional interrelations among its members.

Clearly, the weak programme of the sociology of knowledge analyses only the social factors in Goldman's third sense, and Goldman and Kitcher's models can be said to have extended it so as to show the (positive) role of 'social desires'. Actor Network Theory could be said to attempt to reduce the social to Goldman's fourth sense. The Strong Programme, however, intends to investigate also the social causal chains including social events in Goldman's two first senses. These are social events that have direct causal implications on the content of scientific beliefs. For instance, a combination of a scientist's desire to persuade other scientists using her beliefs about their beliefs cannot but have a causal effect on her production, and the beliefs it will engender in the scientific community. The rhetoric used, such as the metaphors exploited, the cases emphasised, the experiment done, are all directed at persuading a scientific community (Beller, 1999). This orientation of the production impinges on the content of science. There are also, importantly, all the inevitable social transactions (education, readings) that form scientists' opinion. But while sociologists of the Strong Programme have attempted to document and describe these causal relations, Goldman wants to block them with the truism that "one cannot believe a desired conclusion simply by an act of will". This blockage is sustained by Goldman's distinction between social factors and epistemic factors. It seems that, with this distinction, social factors act on will only – one can have social desire, but the social beliefs and social transaction above mentioned are left out: they are damned from the social category, because of their subversive, relativistic, flavour and cannot enter the sacred category of the 'epistemic'. The more comprehensive and correct characterisation of the qualifier 'social' provided by Goldman himself does not warrant the distinction epistemic/social: social events enter the causal chains of scientific knowledge production and their consequences on the content of science cannot be so easily dismissed.

Kitcher and Goldman have had genuine integrative projects: they have integrated, in a first period, cognitive science into philosophy of science, and, in a second period, social studies into philosophy of science. The have developed powerful frameworks for normative epistemology, informed, in turn, by psychology and sociology. They have not, however, integrated cognitive and social studies of science for the development of a science of science. Moreover, I have argued that their (modest but still) foundational goals—showing that science is worthwhile—have led them to indiscriminately reject the great resources of social studies of science. In the end, Kitcher and Goldman's approach furnishes, I believe, an example of integration that does *not* advance towards naturalistic explanations of scientific knowledge production.

2.2.4 PSYCHOLOGY AND THE MYTH OF THE ISOLATED SCIENTIST

Epistemology has, until recently, entertained the myth of the isolated, rational, scientist. This assumption is much present in logical positivism, where science is pictured as the solitary activity of applying inductive principles to one's observation. Cognitive students of science should be careful not to merely translate the principles of logical empiricism in terms of evolutionary biology and cognitive psychology. Pickering (1991) accuses Giere (1988) of doing such a move. Nersessian (1995), answering Shapin's criticism of cognitive studies (1982), asserts on the contrary that "natural rationality" is what is being empirically investigated in cognitive history of science and that it is not assumed to be in the "proper working order" that corresponds to the epistemologists' wishes. This point is important because foundationalist goals and ideas can bias psychological theorising of scientific thinking in favour of the traditional rational isolated scientist picture. For instance, the reply of Giere (1992) to Pickering (1991) unfortunately gives ground to the accusation by reaffirming that:

> the overall project of [my book] was to build a theoretical framework for the study of science somewhere in the middle ground between traditional philosophies of science and the new social constructivist sociologies of science.

The project of resolving the battle between rationalist philosophers and relativist sociologists has generated important work in the last decade in science studies. Works from Longino (2001) or Solomon (2006), among others, try to find some middle way and common grounds between a reformed and modest rationalism and a modest sociologism. These attempts to develop a "picture of science between the extremes of enlightenment rationalism and cultural relativism" (Giere, 1992) are done at the expense of a radically naturalistic science of science: epistemological notions re-enter into the account to cool down the qualms generated by relativism, but hindering a fully causal, mechanistic, explanations (remember that 'he is right' and 'he says the truth' are epistemic evaluation rather than causal accounts). The confusion is even greater when the term 'cognitive' is equated with 'rational' as does Longino (2001): this implies that when causes are cognitive or psychological, one needs not to worry about their rational character. But this is clearly wrong: falsehood does not only come from malevolence and interest in saying false things, it mostly comes from error, from cognitive processes that do not lead to true beliefs, and that do not conform to what should be done for arriving at the truth. A community can always consider some thoughts as not complying with the endowed norms of good reasoning. Here again, integrative projects are deviated towards putting philosophy back again, rather than pursuing naturalisation. Here again, philosophical concerns are foundational, and hinder the development of a science of science that is independent from epistemic evaluations (see section 3.1.1 about the symmetry principle). The problem here is not only the veto put on social studies of science, but also the unrealistic picture of individual's cognitive processes that one may be tempted to develop. Cognitive scientists of science has been seen as merely transferring the positivists' foundational logic and its purported virtue within the heads of scientists. Thus, citetwoolgar89 characterizes the cognitivist stance as "the idea that mental or other inner processes enable the *rightful* perception of an already existing world" (p. 206; my emphasis). The association of cognitive studies with a psychologised logical positivism has justified reactions against cognitive studies resulting in social reductionism. The association was also maintained because 'cognitive' is, unfortunately, often used interchangeably with 'rational' (see, e.g. Longino, 2001).

Nersessian (2005) provides supplementary, but related, reasons why integration has been hampered, which regards theories of cognition. She argues that the perceived divide between socio-cultural and cognitive accounts of science and technology arose from the implicit and explicit notions of "cognition" used by the protagonists. These notions have motivated reductionist attitudes from both sociologists and cognitive psychologists. She says:

> Implicit echoes of Cartesian dualism underlie the anticognitive stance in sociocultural studies, leading to sociocultural reductionism. On this side, Cartesianism is rejected as untenable but, rather than developing an alternative theory to encompass cognitive explanatory factors, these are rejected outright. Within cognitive studies [...] Cognitive reductionism identifies cognition with symbol processing that, in humans, takes place within an individual mind. (p. 18)

Early cognitive studies of science have developed discovery programmes as models of cognitive heuristics that would be used by scientists (Langley et al., 1981). The cognitive reductive view of science would present these heuristics as *the* principles out of which science evolves. Sociologists of have strongly reacted against this individualistic view: much of the constructive processes of knowledge developments are indeed missing from such accounts. Two of the much-criticised Cartesian characterisations are: (1) the alleged independence of cognitive processes from the medium in which it is implemented, and (2) the treatment of the socio-cultural dimension, reduced to socio-cultural knowledge inside the mind of individuals.

The reason for Latour & Woolgar's call for a moratorium of cognitive studies of science (Latour and Woolgar, 1986; Latour, 1987) is the partially justified association of an erroneous Cartesianism with cognitive studies of science. The criticism, she continues, would then bear on a cognitive reductivism in which "the social and the cultural environments [...] are treated as abstract content on which cognitive processes operate." Sociologists' reaction against Cartesian and individualist view of cognition has been to "throwing out the baby with the bath water," and go for social reductionism. Nersessian argues that this reaction is misguided, as recent developments in cognitive science are deeply aware of the deadlocks of the Cartesian view of the mind. Nersessian appeals to theories that develop "environmental perspective on cognition," which provide satisfactory answers to the criticism of the Cartesian view of cognition. She argues that these approaches can eventually be used for integrating cognitive and social studies of science (see section 2.3.2 below).

2.3 Integrative projects from the psychology of science

Although cognitive science actually started on an individualistic basis, caricatured as the Cartesian 'brain in a vat' picture of cognition, integrative projects have mostly originated in psychology of science, rather than in sociology of science.

Feist and Gorman (1998) list four subfields in the psychology of science:

developmental psychology of science analyses how and why certain individuals become scientists, and the origin and the developments of the skills and abilities involved in doing science.

personality psychology of science investigates personality differences between scientists and non scientists, and between eminent and less eminent scientists. It also investigates the relation between personality, and theoretical persuasions and scientific behaviour. Its results are, for instance, that the scientists are generally more conscientious, dominant, and emotionally stable than non-scientists.

social psychology of science investigates how the thoughts, feelings and behaviour of scientists are influenced by the actual, imagined or implied presence of others. Feist and Gorman complains, in their 1998 article, that the field is still under-developed.

cognitive psychology of science analyse how information is treated and transformed, through which processes and thanks to which abilities, so as to output scientific representations.

These studies have been considering seriously social phenomena. For instance, developmental psychology of science is interested in the social conditions where education takes place, and it deals with socially relevant question such as the one regarding the relation between gender or age and mathematical abilities. Personality psychology has also connections with the sociology of science, since personality determines both social behaviour and theoretical choices. Social psychology is tautologically connected to sociology. Cognitive psychology is more and more turning towards analyses of the social determination of cognition and its social embodiment.

It is possible to distinguish two (ideal-) types of contributions of psychology to the understanding of science as a natural phenomenon. There are, one the one hand, studies that analyse factors hindering or enabling the development of science, as with the study of what makes scientists be creative; and, on the other hand, studies that analyse the very processes through which scientific knowledge is produced, as with the study of the analogical reasoning processes used by scientists. The first type of contribution explain the psychological conditions of scientific knowledge production, while the second type of contribution explain why science come to have the content it has. To a large extent, personality psychology develops contributions of the first type, and cognitive psychology of the second type. In sociology of science, those two types of contributions have been distinguished because traditional, Mertonian, sociology of science was willingly limiting its contribution to the first type. Bloor (1976) coined this research, the 'weak programme' in the sociology of science and advocated a 'strong programme.' The idea beyond this terminology is not at all to dismiss traditional Mertonian sociology, but to reject the limits set on sociological inquiry. While the ambition of a strong programme in the sociology of science has been strongly attacked, there has been no such reaction against a similar strong programme in the psychology of science. This is good news for the integrative project: the accepted strong programme in the psychology of science could ground and justify the controversial strong programme in the sociology of science, thus issuing and integrated strong programme in the science of science. This is, in any case, the ambition I want to pursue. This is why I focus on inquiries of the thoughts and cognitive processes of scientists, and leave aside the interesting research on the psychological facts that make people more or less able to contribute to the scientific enterprise.

I will mention two trends in cognitive psychology of science, which have gone towards integrating the analysis of social phenomena involved in knowledge production. These trends have been best described by Nersessian, in her 1995's article *Opening the Black Box: Cognitive Science and History of Science*, and in her 2006's book chapter *Interpreting Scientific and Engineering Practices: Integrating the Cognitive, Social, and Cultural Dimensions*.

2.3.1 Cognitive history of science

Nersessian (1995) introduces the methodology and prospects of cognitive history of science as a subfield of cognitive science, which studies "the "thinking practices" through which scientists create, change, and communicate their representations of nature [3]" (p. 194). Cognitive history of science reconstructs historical events that are amenable to cognitive analysis. Work in cognitive history of science include analyses of the works of Faraday (Tweney, 1985, 1991; Gooding, 1990,?), Maxwell (Nersessian, 1984, 1992b, 2002c), and Bell and Edison (Gorman, 1992). These works have attempted to reconstruct scientists' thoughts and activities with cognitive notions such as "schemata," "mental models," "heuristics," and "procedural knowledge." They have analysed through which cognitive processes innovation and conceptual change arise in real historical cases. The study of these historical cognitive events can contribute to the understanding of cognition, as it provides new data on cognition done in non-experimental settings. Instead of results of tasks performed in the psychologists' laboratory, cognitive history of science analyse how scientific cognition is done in real cases. However, this analysis is being informed by results from experimental psychology. Nersessian criticises previous integrative attempts in the history of science, which were fitting the history of science to models imported from psychology. Piaget, for instance, fitted the history of science into the framework of his theory of cognitive development (p. 197).

[3] Presumably, the characterisation can be adapted for mathematics by talking of mathematicians and mathematical representations

There should be, as Nersessian (1995) puts it, a virtuous circle where some assumptions from cognitive science are accorded privileged status in order to get the historical analysis "off the ground," but could further on be subject to critical scrutiny Corrective insights should move in both directions: from cognitive science to cognitive history and the reverse. All authors in cognitive history of science have advocated using a plurality of methods for investigating science as a natural phenomenon. They also have adopted a non-reductivist view of the cognitive approach, and assumed that "science is one product of the interaction of the human mind with the world and with other humans" (p. 195).

Cognitive history of science analyse scientific cognition as it occurs in historical contexts. Scientists are socially situated, and this situation accounts in part for their thoughts and behaviour. There is therefore a bridge built between social and cognitive studies of science that cannot be found in, e.g. Simon's work It is this direction that I will pursue. The limits of these work is that they say little on the historical development of science. Although historical, these works remain synchronic in the sense that they study one individual in his relatively fixed historical context: the cognitive evolution studied is limited to individual scientist history. These studies thus provide a relatively static picture of a moment in the history of science. A further integrative step is needed, from cognitive to social history of science, in order to restituate the cognitive production of individual scientists as participating to the development of science, and understand the causal effects that their work have had on scientific cultures. In the epidemiological framework I will advocate in the next part, the object study is the evolution of the distribution of scientific representations in the scientific community—a historical and social phenomenon that is constitutive of science. Representations are produced, computed and transformed in different media—including scientists' brains—and through time. In the causal chains that issue a distribution of scientific representations among a community, the thoughts and deeds of some eminent scientist is but one set of events, and the reasons of their causal effects can be questioned. Cognitive history can aim, eventually, at specifying the psychological foundations of the historicity/evolution of science, and how psychological factors determine the content of science The second part of this thesis is a stab in that direction.

2.3.2 Environmental approaches in cognitive studies of science

Nersessian (2006) presents new environmental approaches in cognitive science as means to renew the work in cognitive studies of science and integrate it with social studies. The environmental approaches argue that studies of cognition need to take the role of the environment into account, not only as an external provider of input to cognitive processes, but also as having a role in cognitive processing. Studies of cognition in authentic contexts of human activity have shown that cognitive processes need be treated as strongly dependant on the contexts and activities in which cognition occurs. For instance, Lave (1988) shows that people perform better when solving mathematical problems in supermarkets than when solving similar problems in tests, because resources and possible actions upon them are different in the two situations. The environment shape and participate to cognition. The use of artefacts in solving cognitive tasks is a case in point: for instance, adding large number is done by writing down material representations of numbers; one uses external representations for computing the results and follows a detailed procedure in adding each row (McClelland et al., 1986, p.44-48). One common point of the environmental approaches is their rejection of the individualistic assumptions and disembodied view of cognition that GOFAI (Good Old Fashion Artificial Intelligence) has developed. Cognition, the environmental approaches argue, does not reside solely in the head; it includes relations and interactions between individuals and their particular situations. Hutchins argues that important cognitive properties that are traditionally attributed to the functioning of the mind are in fact properties of larger cognitive systems that can include cognitive artefacts, and several human agents.

Nersessian concludes her review of the field with the promise that these approaches "offer a substantially new way [...] of thinking about the social-cognitive-cultural nexus in science and engineering practices." As a matter of fact, cognitive history of science, which studies scientific cognition in context, participates to the study of the role of the environment in cognition, and it confirms the needs for environmental approaches of cognition. Environmental approaches, however, especially draw attention on the social organisation of cognition, on its embodiment in external artefacts, and on the role of action in cognition. Works adopting environmental approaches in cognitive studies of science include numerous works that show the role of the experimental apparatus, and its manipulation, of cognitive tools (e.g. computers), and of external representational media in the production of scientific knowledge. For instance, the role of diagrams and pictures in scientific cognition has recently been the focus of much interest (Gooding, 2004, 2005; Roth and Bowen, 2003; Roth, 2004, e.g.). To the extent that environmental approaches analyse events that happens outside of the brain and which contribute to the production of knowledge, many studies in social studies of science contribute to the cognitive study of the environmental aspects of cognition. The practice turn in science studies parallels the move towards situated cognition in cognitive science, and focuses upon what is done rather than what is thought. As with cognitive history, one challenge is to understand the evolution of science on the basis of these new analyses. If cognition is situated and distributed, then cognition can evolves as the situation and the distribution change. We can then see that scientific cognition evolves, and this evolution boosts changes in scientific knowledge itself. But how do changes in cognition occur in the evolution of science? I attempt to answer this question in the third part of the thesis. One thing that the answer involves is a return to the mental processes, which are sometimes forgotten due to the enthusiasm for the analysis of practices, and the role of the environment in the making of scientific knowledge.

2.3.3 The psychology of social studies

Psychological theories that are the most likely to integrate social studies are the ones that come from social studies themselves. These theories, however, are rarely explicated—suggesting that most social science is based on folk psychological theories. There is, however, a set of assumptions that are shared by most social scientists about the mind as vector of cultural knowledge: it is seen as a blank slate upon which culture puts its stamp. One of the most explicit attempts to introduce psychological theory into the social science is due to Bourdieu (1977), who developed the notion of the *habitus*. The habitus is a system of dispositions that agent acquire during their interaction with their objective social conditions. It is a structure internal to the agent, realised in her body and mind, and that account for her behaviour. The habitus, as acquired schemes of perception, thought and action, is a psychological notion. Its theory is a means to account for the relation between social objective structures and the subjective experience of the agents. The objective social structures are inculcated into the agents via the habitus, who then reproduce the structure via their behaviour and practices.

Bourdieu has applied his theory to the 'academic world' (1975; 1976; 1984; 2001), thus contributing to science studies. His theory of the habitus, here, evokes Durkheim's Kantian programme: uncovering the categories of thoughts and their social origins. Bourdieu talks of the "academic transcendental" in order to designate the categories of professorial understanding as, e.g., the classificatory schemata that French teachers implement in assessing students. To a large extent, Bourdieu's analyses bear more on the processes of social reproduction in the academic world, than directly on the processes of knowledge production. My concern, however, is about the validity of the psychological theories developed by Bourdieu. Is it sufficiently informed by *current* theories in cognitive psychology? Lizardo (2004) analyses the sources of Bourdieu's notion of Habitus, and emphasises the influence of Piaget's work. Bourdieu's theory of habitus is not a mere product of folk psychology, but yet seems especially fitted to the sociological theory of Bourdieu. As focused on social reproduction, it gives little place to the psychological mechanisms of creativity, and it does not recognise the important role that non-social determinations of cognition actually have. The challenge, therefore, is to provide social theory with more comprehensive and updated psychological theories, which come from directly from psychologists and cognitive scientists themselves. The idea is that it is safer to ask psychologists and cognitive scientists about the psychological constitution of social agents, than to make up one's own purpose oriented psychological theory. It is safer not to ignore the work of those who have taken the mind as their object study—all the more so that, as mentioned above, cognitive scientists have come to realise the importance of the social context in cognition. In the next part, I will show the relevance of findings and theories in developmental psychology, neuropsychology, cognitive psychology and evolutionary psychology, for the study of scientific evolution as a socio-historical phenomenon.

2.4 WHY THERE SHOULD BE A COGNITIVE ANTHROPOLOGY OF SCIENCE

The assumption that scientific evolution is a socio-historical phenomenon implies that the sociological study of science is primary and essential. If cognitive studies want to contribute to the study of scientific evolution, then they have to show their relevance for the study of this social phenomenon. Cognitive history of science and studies of situated and distributed cognition already show the relevance of the cognitive approach for the social study of science. In this section, I will show that cognitive anthropology comes up with further prospects. It addresses relevant question and provides good means for the integration of social and cognitive studies of science. As opposed to the homo-economicus, the agent of cognitive anthropology is understood as being endowed with rich resources for understanding and interpreting the world. This section shows that the theories and methods of cognitive anthropology are particularly adequate for the study of science.

'Cognitive' as a naturalised notion

Anthropology has long been struggling with issues related to cognition and rationality. Since Levy-Bruhl's distinction between primitive "pre-logical" thought and rational modern thought, Rationality has often been challenged as a mere Western ethnocentric presupposition (Wilson, 1970). Cognitive anthropology allows restating the problem anew by analyzing cognition not as thought processes leading to true beliefs (in normal conditions) but as mental mechanisms or properties sustaining the many diverse cultures. In that perspective, cognitive anthropology of science will not aim at discovering the essence of science; it will rather investigate the ways in which the mental apparatus allows the production of the cultural phenomena found in the history of science, and reciprocally how the specific cultural environments of science constrain or inform mental processes.

The interplay between scientific cognition and scientific cultures

With a cognitive anthropological approach, one can investigate *empirically* the characteristics of scientific thoughts and practices. Being freed from a normative agenda or its associated essentialist claims about the nature of science, one can aim at a description of scientific cultures, and hope to find out its most salient traits. But if one can identify scientific communities, if only as self proclaimed communities that work at specifying their identities and maintaining cultural boundaries through differentiation (Ellen, 2004), it is far more difficult to characterise these cultures in terms of cognitive practices. Again, the question is a classic one in cognitive anthropology: do different cultures imply different ways of thinking? The most radical answer can be found in strong *cognitive* relativism, according to which cultures do have a very important impact in framing mental processes and abilities. The Sapir-Worf hypothesis, for instance, hypothesises that thinking and perceiving radically differ from one language community to another. Most recent findings, however, have shown that the mind is richly endowed with innate structures and abilities that enable and strongly constrain human thinking (see next chapter for a more complete account). Berlin and Kay's study (1969), for instance, suggests that the extensions of colour terms are constrained by universal cognitive and biological factors rather than being totally relative to cultural contingencies that would constrain perception. There is, in cognitive anthropology, a research trend that aims at articulating the causal relation from the innate cognitive constraints of our mental apparatus to the diversity of cultural phenomena (Barkow et al., 1992; Hirschfeld and Gelman, 1994; Sperber, 1996a). Applying this research programme to the sciences of life led Atran (1990) to hypothesize the existence of an innate ability to reason about living kinds that has constrained the history of Natural History as well as current research practices in neo-darwinist theory.

The continuity hypothesis

Either the mind/brain is sufficiently plastic to allow drastic changes in the organisation of the mental apparatus during the course of scientific education (Churchland, 1988), or lay and scientific mental processes remain mainly similar, which is the continuity hypothesis. The question provides a central research direction in cognitive anthropology of science and sends us back to much other research: Neurobiology on the plasticity of the brain, developmental psychology on the impact of scientific education, evolutionary psychology with regard to the plausibility of the abilities hypothesized, comparative and historical anthropology, etc. In cognitive anthropology of science the issues concern the relationships between folk theories and scientific knowledge and practices. Do cognitive dispositions afford and constrain science in the same way as they afford and constrain folk knowledge? Can the development of science be seen as a cultural process of emancipation from cognitive constraints? Does science manage cognitive resources such as memory, imagination and reasoning abilities in the same way as other cultural institutions such as religion? The study of apprenticeship and enculturation during scientific education allow us to understand how thought processes are oriented, framed or made possible so as to engender specifically scientific thinking (Roth, 2004; Alac and Hutchins, 2004; Poling and Evans, 2004; Kurz-Milcke et al., 2004). There are many ways to understand the continuity hypothesis, depending on which cognitive phenomena are taken to be stable across cultures and life span, or which analogy one wants to draw between scientific and lay cognition (see Carruthers et al., 2002; Erana and Martinez, 2004).

Mental and cultural models

Mental models are mental constructs that represent a situation, event or process and have structural similarities with what they represent (Johnson-Laird, 1983). They are studied both in cognitive studies of science and in cognitive anthropology, but with a rather different perspective. In cognitive studies of science, the research on mental models was initiated as a reaction to the shortcomings of logical positivism, which aimed at describing scientific knowledge in terms of axiomatic systems, and scientific reasoning in terms of logical, syntactic, operations on propositional representations. Contrary to this view, mental models have been shown to play a major role in the 'cognitive structure of scientific theories' (Giere, 1988, 1994) and in scientific reasoning – such as analogical, visual and simulative modelling (e.g. Nersessian, 1992a; Magnani and Nersessian, 2002). In cognitive anthropology, the emphasis has been on cultural models, which are mental models that are culturally shared, such as the American model of marriage (Quinn, 1987). But in science also, a mental model acquires a significant role only when it is shared by a scientific community. In mathematics, for instance, proof methods, argumentative methods and proof strategies (Bendegem and Kerkhove, 2004)appear to be cultural models for mathematicians in much the same way as Quinn's model of marriage for Americans. While the importance of the distribution of a scientific mental model among the scientific community shows well the usefulness of the cognitive anthropological approach, there is also another sense in which scientific models are culturally shared and consisting of social phenomena: Scientific models do not only take place in scientists' minds, they also often have an institutional, material and social reality. This is because of the embodied, situated and distributed aspects of (scientific) cognition.

Embodied, situated and distributed scientific cognition

The environmental approaches in cognitive science partly originate from the work of cognitive anthropologists (Lave, 1988; Hutchins, 1995, e.g.). Cognitive anthropology is well equipped for the study of situated and socially embodied cognition, since it studies cognitive events in their natural context. Hutchins (1995) has advocated doing "cognitive ethnography" in order to describe how humans interact with the environment when solving cognitive tasks. When the flow and transformations of scientific representations are studied without an arbitrary restriction to purely mental processes, one can take on the task of describing cognitive systems that are distributed among human agents and artefacts. The ensuing framework leads to the analysis of the social organisation of the cognitive systems that produce scientific knowledge, and points out the essential phenomena 'where the cognitive and the social merge' (Giere and Moffatt, 2003). What are the specific cognitive architectures of scientific institutions? How do these structures relate to the production of scientific knowledge?

The epidemiology of scientific representations

One way to describe the evolution of science as a social phenomenon is to view it as changes in the distribution of scientific representations in the scientific community and its environment. At one point in history, representations of the Ptolemaic system were transmitted and used to account for the movement of the planets; at a later time heliocentric representations of the movement of the planets were transmitted and used. At one point in history, Ptolemaic representations were in the head and embodied in the working tools of astronomers; at a later point it is heliocentric representations that were distributed among astronomers and their working environment. How and why did this change happen? Sperber's epidemiology of representations is a framework that frames questions about culture in the above way, and attempts to answer by specifying the factors and mechanisms of the distribution of representation. Sperber (1996b) presents the epidemiology of representations as follow:

> A human population is inhabited by a much wider population of mental representations. The common environment of that population is furnished with the public productions of its members, some long lasting, like buildings, other ephemeral, like the sounds of speech. Particularly important among these productions are (tokens of) public representations. Typically, productions have mental representations among their causes, and mental representations have productions (in particular public representations) among their causes. There are thus complex causal chains where mental representations and public productions alternate. In many cases, representations (mental or public), occuring in these causal chains inherit some of the semantic properties of the representations (mental or public) of which they are causal descendants. A variety of inter individual processes bring about this match between causal and semantic relationships. Processes of imitation and communication can be described as having the function of bringing about such semantic similarity.

Applying this framework to science means tracking down scientific representations, public or mental, and reconstructing the processes that produced, transformed and distributed them; it means reconstructing the causal chains with alternating mental and public representations, with events such as thinking and communicating, and that eventually changed the distribution of scientific representations. Together with historical causal reconstructions, a theory of the mechanisms and factors of distribution of scientific representations can be elaborated — with the usual feedback loop between theory as enabling the interpretation of historical data, and historical data used for enriching or challenging the theory.

In this chapter, I have reviewed attempts from philosophers and cognitive scientists to integrate cognitive and social studies of science. While some attempts are not aiming at the development of a science of science and can be misguiding for the descriptive and explanatory enterprise, other attempts are showing nice prospects, which I will attempt to purse. My own focus concerns the evolution of science as a cultural phenomenon; with the persuasion that it importantly involves mental processes. I eventually pointed out at the rich resources that cognitive anthropology can bring to science studies. The cognitive anthropology of science is at the crossroads of several rapidly developing disciplines: cognitive science, which increasingly provides tools for the study of scientific thinking; science studies, and in particular the anthropology of science, which is enriching the subject with numerous case studies; naturalised epistemology, which is constantly reworking its philosophical assumptions thus opening new directions for the naturalist study of science; and finally, cognitive anthropology and ethnoscience, which make a valuable contribution in terms of theory, methods and empirical data. Thus, the cognitive anthropology of science benefits from several paradigms, traditions and research methods. First, cognitive anthropologists can show how cognitive constraints have contributed, together with historical and cultural factors, to the contents of a given science. Second, sciences can be analysed as specific cultural models or schemas that frame individuals' cognition. Scientists at work, and, more controversially, people in their everyday activities, appeal to specific ways of thinking informed by the 'culture of science.' Third, sciences are cultural objects of particular relevance for the cross-cultural study of notions such as truth or causality, and cognitive operations such as reasoning or categorising. Fourth, scientific practice can be analysed as cognition distributed among scientists and scientific instruments. Identifying scientific cultures is often easy, because of their self proclaimed constitution. But determining and describing their salient social and cognitive traits is something much harder that requires detailed empirical investigation. Cognitive anthropology, as the study of thought in cultural context, but also as the study of culture as constituted through people's thinking and interacting, should therefore enable stressing the social or the cognitive when needed – i.e. when the determinants or causal factors of the scientific event to be explained are actually (but non exclusively) social or cognitive. The epidemiology of representations is the framework, drawn from cognitive anthropology, that I will use, and advocate using, along the thesis. I take it that one goal of naturalised epistemology is to account for scientific knowledge production with a description of the causal chains that produce and distribute scientific representations, thus issuing science—seen as a cultural and cognitive object. These causal chains include both mental, and social and historical events. In the next chapter, I will review to which extent the sociology of scientific knowledge can provide good methodological and theoretical bases for the development of integrated, naturalistic, science studies.

Chapter 3

Integrating the Strong Programme

This chapter investigates the role of social studies of science in an integrated science of science. It includes an appraisal of the methodologies developed in social studies of science based on an analysis of their openness to, and compatibility with, psychological and cognitive studies. Is it possible to give important explanatory roles to social studies of science without embracing social reductivism? Sociological research that restrain inquiries to the institutional framework of science and social determinations of the rate of growth and direction of scientific research—coined the 'weak programme' by Bloor—is broadly accepted as valid and useful. But the research initiated by sociologists of the Strong Programme holds that sociology is essential for explaining the making of scientific knowledge: not only are social phenomena constitutive of science, they also determine its very content. It is often believed that these developments in the sociology of scientific knowledge are reductionists. But it is not necessarily so. The claim that sociology is relevant for explaining more things than previously thought does not imply that other fields loose of their relevance. Psychology, in particular, can always participate to the explanation of social phenomena, since social phenomena involve people whose actions are determined by brain/psychological processes. In that perspective, claiming that the making of science is social through and out implies that *psychology* is relevant for explaining it.

Most authors in social studies of science have not thought this way, and have neglected cognitive studies of science, or even attempted to debunk the whole enterprise. I argue that the methodology and theories of the Strong Programme, among current theories in the sociology of science, are the best candidate for integration. I develop on the compatibility claims made by its proponents and review the arguments about the essential role that sociology should have in a science of science. Among the most controversial claims of the Strong Programme is the assertion that social phenomena do have a causal role in the formation of scientific beliefs. Yet, belief formation is usually thought of as a psychological phenomenon. How does this claim impart analytical tasks between cognitive and social studies of science?

In the first section, I give a brief presentation of the Strong Programme and point out at its positive attitude towards psychology and cognitive science. In the second section, I compare the Strong Programme to other methodological proposals in social studies of science with respect to their ability to integrate information from cognitive psychology. In the third section, I show how the theses of the Strong Programme could actually integrate psychological assertions.

3.1 THE STRONG PROGRAMME AND ITS ATTITUDE TOWARDS PSYCHOLOGY

The Strong Programme developed in the 70's with the intent to further the naturalisation of epistemology. This meant, for the protagonists of the Strong Programme, bringing the expertise of empirical sciences to bear on the study of phenomena traditionally studied with philosophical, analytical, means, and with foundational goals. The empirical science they advocated as highly relevant was sociology. The principles of the Strong Programme are mostly methodological, specifying why and how sociology can be useful for the naturalistic study of science. Sometimes, these methodological points pass through direct attacks against rational reconstruction, presented as an inappropriate method hindering causal explanations of the development of scientific knowledge.

Bloor (1976, p. 7) provides the initial characterisation of the Strong Programme as research complying with the following four tenets:

> 1. It would be causal, that is, concerned with the conditions which bring about belief or states of knowledge. Naturally there will be other types of causes apart from social ones which will cooperate in bringing about belief.
> 2. It would be impartial with respect to truth and falsity, rationality or irrationality, success or failure. Both sides of these dichotomies will require explanation.
> 3. It would be symmetrical in its style of explanation. The same types of cause would explain, say, true and false beliefs.
> 4. It would be reflexive. In principle its patterns of explanation would have to be applicable to sociology itself.

The first important thing to note for our concern is that the Strong Programme is not social reductionist. This is stated right in the first tenet of its characterisation "there will be other types of causes apart from social ones." The first tenet already calls for integration with other scientific approaches in science study, as the different types of cause interrelate in causal chains that issue "distributions of belief" (p. 5); so the different disciplinary studies of each types of causes should collaborate to provide comprehensive explanations. The two next tenets constitute the symmetry principle, which constitutes the basic and most controversial claim of the Strong Programme, and its departure from the 'weak programme' in the sociology of science.

3.1.1 The symmetry principle

It is intended at insuring that explanations will be causal through and through. It enjoins the scientist of science to seek the same types of causes for both true and false, rational and irrational, beliefs. What does Bloor means by 'same types' in "same types of causes"? Explicitly, the types referred to are not restricted to social types. Sociologists of the Strong Programme have indeed numerously lamented this misunderstanding and insisted, in particular, that input from the empirical world is a genuine and important cause in the processes of belief formation. When would causes be of different types, then, for the Strong Programme? Here is an example of asymmetric explanation: scientist A has false beliefs because cause x made him depart from rational thinking; scientist B has true beliefs because they were caused by the nature and structure of the phenomena he investigated. There are two types of causes: causes of deviance from some norms of rationality—presumably social causes—, and causes of true beliefs found in the natural world itself. The problem with such asymmetric explanations is that it is uninformative with regard to the causes of true beliefs—it merely say that a scientific holds a true belief because it is true—and in the danger of being anachronistic with regard to the causes of false belief—deviance is characterised on the basis of an evaluation that can involve norms, beliefs and terms that were not present at the time when the belief was formed. The main target of the symmetry principle is rational reconstruction, which explains how rational those holding true beliefs have been, and why those holding false beliefs have been irrational. The symmetry principle advocates explaining beliefs independently of how these beliefs are evaluated. It denies that sociologists of science should restrict their investigations to specifying which social conditions lead to 'scientific' behaviour and which do not. Indeed, one result of the application of the symmetry principle, and the one mostly used and appealed to by the sociologists of the Strong Programme, is that social factors have a causal role in the formation of both true and false scientific beliefs. One essential reason for this, is that a belief is scientific only when it is deemed to be so by the scientific community—a social phenomenon by itself. The claim that social phenomena are pervasive causes in scientific knowledge production is enabled by the symmetry principle; however, the principle does not ban causal psychological explanations:

> The symmetry requirement is meant to stop the intrusion of a non-naturalistic notion of reason into the causal story. It is not designed to exclude an appropriately naturalistic construal of reason, whether this be psychological or sociological. (Bloor, 1976, afterword to the second edition)

How is it possible to comply with the symmetry principle? How can we explain the formation of true belief if not with the facts that make these beliefs true, since it is these facts that are the causes of the beliefs? As the input from the external world is recognised to be a cause in the processes of belief formation, a description of this input should figure in the explanation of the process. But the best description of it is admittedly the one given by the scientists themselves. How can the analyst go further than "A believes that p because p"? Saying that Millikan believed that the charge of the electron is 4.7×10^{-10} esu because the charge of the electron is 4.7×10^{-10} esu is not very explanatory. The strategy advocated by the symmetry principle is to bracket off one's belief in the truth of the scientist's claim under investigation, and take it only as an object of study. The historian needs to go back to what Milllikan actually perceived — to his experimental setting and its observations. Evidence that motivated scientists to hold some beliefs should figure in the explanation of why they hold the beliefs. However, some further explanation is needed regarding why they took the evidences as being evidence for the beliefs the hold. It is this status of evidence that is not taken for granted, and the processes that lead scientists to provide it in need of explanation.

3.1.2 METHODOLOGICAL RELATIVISM

The method of bracketing the truth of *some* claims when explaining why they were made constitutes the Strong Programme's methodological relativism. When studying a scientific controversy, for instance, sociologists of the Strong Programme abstain from appealing to the fact that one party has true scientific beliefs while the other hold false beliefs. This epistemic evaluation should not figure into the causal explanation of the advent of the belief because the truth of a claim has by itself no causal power — it is an abstract property of the claim.

What is the relation between methodological relativism and other types of relativism? Bricmont and Sokal (2001) assert that "methodological relativism makes sense only if one adheres to cognitive relativism", where cognitive relativism is defined as the thesis that "truth or falsity of a statement is relative to an individual or a social group". Their argument is worth a brief account because it includes two widespread misunderstanding. It goes as follow: although sociologists of scientific knowledge adopting methodological relativism have scientistic claims about their social analyses, which they want to be causal and materialist, they analyse results of the natural sciences independently of their truth or falsity. This method, they say, makes sense only if scientific knowledge is indeed independent from its truth and falsity. This implies that the sociology of science, which has not a better scientific record than the natural sciences, is also independent from its truth and falsity. The way out of this embarrassing conclusion, say Brickmont and Sokal, is for sociologists of science to adopt the idea that no theory is objectively better than another – i.e. cognitive relativism. The misunderstanding are: first, the sociology of scientific knowledge attempts to give causal accounts of knowledge claims. Its object study is the behaviour of scientists, and it is not to be confused with an evaluation of the truth of the content of their claims. The evaluative project belongs to epistemology, not to sociology, which intends to be descriptive. The second misunderstanding concerns the method of bracketing the truth of some statements in order to give a causal account of the events that lead scientists to claim that these statements are true. The bracketing method is not to be confused with a general scepticism. It is a methodological principle justified by the fact that truth has no causal power by itself. However, methodological relativism is keen to use scientific knowledge in general for the purpose of its naturalistic enquiry: in particular, the knowledge of sociologists is not bracketed out, and sociologists of science take position on the truth of some sociological claims. As a conclusion, methodological relativism is fully compatible with the idea that scientific knowledge is indeed providing important knowledge about the world, or that it is worthy for any other reasons (e.g. pragmatic reasons).

Methodological relativism, I repeat, does not deny the role of the outside world in belief formation, nor does it deny that there are cognitive processes. Bloor (2004b) specifies its intention as follow:

> For the purpose of the sociology of knowledge, relativism is the thesis that the credibility of all beliefs calls for explanation in terms of local, contingent causes.

Here is a second example of bracketing the truth of scientific claims for the purpose of analysis, which is used, in Spranzi's analysis (2004), to track down the cognitive events issuing scientific beliefs, viz. the belief that there are mountains on the moon. Spranzi says that Galileo saw dark marks on the moon, rather than shadows; she is putting into brackets her belief that the dark marks are caused by the shadows of mountains of the moon. This bracketing allows her to explain why and how these marks came to be 'seen' as shadows of mountains. An analysis of the flow of information through cognitive analysis allows identifying the input and the output of cognitive systems: by recognising the causal role of the former in the production of the latter, one can point out where the external world (i.e. external to the cognitive system) intervenes. But of most interest for the scientist of science is what happens in between the input and the output: the social and cognitive construction of scientific knowledge.

3.1.3 Social theory of the Strong Progamme

Although the methodological points listed above—naturalisation of epistemology, commitment to causal explanations, the symmetry principle and methodological relativism—form the main bulk of the Strong Programme and what has been taken by others as characterising it, sociologists of the Strong Programme have also developed social theories, which they have applied to the social study of science. These are mainly *meaning finitism*, *a theory of social institutions*, *interest theory*.

Meaning finitism is first a logical claim restating the problem of induction in the case of word application. It is the denial that words have intrinsic power that determine their own application. It relies on the fact that a finite number of applications of a word does not determine, on its own, future applications of that term. In particular, a word is applied to two different items in virtue of a something that they have in common; there is a similarity between the two items. However, anything can be found to resemble, in some way or another, to anything else. So resemblance is not sufficient for determining word application. In the face of this under-determination problem, one can call on the use of words in network of beliefs, which determine future word applications. For instance, knowing that ducks are web-footed may determine one's future application of the word 'duck.' However, definitions and characterisations rely, at some point, on ostensive learning in order to be understood; and ostensive learning is based on the observation of a finite number of word application. The consequence is that future applications of words are open ended. Moreover, every act of classification can be qualified as incorrect and can be revised. The consequence that sociologists of the Strong Programme derive from these observations is that there are local, contingent facts that determine word application. Every act of word application is sociologically problematic, whether it is taken for granted and based on feeling of sameness and individual's inductive propensity, or the topic of controversies. Word application can be correct or incorrect, and this depends from standards that are sufficiently consensual to be invoked in argumentation. Thus, meaning finitism is a semantic theory that draws on holism, pointing out the inter-dependency of word applications, and social externalism, pointing out the social aspects of epistemic judgements in word application.

The Strong Programme's theory of social institutions is well developed by Barnes and Bloor. Barnes (1983) developed an early picture of institutions, as the product of individuals referring to both facts independent of the institution, and to facts arising from the general presumption that the institution exists. For instance, using currencies implies referring to the material implementation of the currency and to the facts that the currency is given value by others. Barnes points out the bootstrapping process involved in the maintenance of institution: people refer to their own practices, and by doing so, enforce and maintain the practices; this is what makes the institution. Bloor has also importantly contributed to the development of the theory, especially by bringing forward Wittgenstein's philosophy on rule application and by showing the pervasiveness of the bootstrapping process just mentioned (Bloor, 1997a). The theory of institution is applied to scientific knowledge, which is shown to have the features of institutions. Scientists, indeed, always appeal to, and rely on, previous conventions for describing and explaining new phenomena. Scientific thinking and acting is done with the help of previously developed scientific beliefs, thus reinforcing the institutional status of the scientific beliefs. Sociologists of the Strong Programme aim at tracking down this conventional and often taken for granted assumptions at the basis of scientists' behaviour. The existence and processes for maintaining or debunking these conventions are social phenomena that the Strong Programme shows to be always present in the making of science.

Interest theory has been much decried by the opponents of the Sociology of Scientific Knowledge. It has been caricaturised as meaning that scientists act only in order to further their own social well-being. As a counter argument, it has been pointed out that scientists have "epistemic interests," interest in the pursuit of truth, that differ from social interests (e.g. Giere, 1994). Sociologists of the Strong Programme deny that one can tell apart epistemic interests from social interests: the making of scientific knowledge is always both epistemic and social. The interests the Strong Programme appeals upon are the causes of actions, which are motivated and oriented towards the achievement of goals. Interest theory is nothing but a general theory of action applied in the context of knowledge making. It is through the activity of scientists that science is made; so science studies must include analyses of the scientists' actions that contribute to the development of science. However, rather than appealing to the traditional belief-desire explanation of action, with its assumptions on human rationality, sociologists of the Strong Programme talk about goal oriented action and aims at developing causal explanations of human behaviour. Why did a scientists did so and so? Let us look, says interest theory, at their interests and goals in doing it. Indeed, 'desinterested' scientific research does not exist, as scientific actions always have causes; in particular, 'rationality' has by itself no causal power and does talk through the mouth of scientists. Thus:

> Goals and interests are associated with scientific research in all actual situations, and operate causes of the actions or series of actions which constitute the research. (Barnes et al., 1996, p. 120)

In the context of science study, interest theory develops into a pragmatic theory of the evolution of scientific knowledge. Change in knowledge, as the outcome of scientists' interested actions, increase "utility of particular cultural resources for particular kinds of prediction and control as described in a particular form of shared discourse" (Barnes et al., 1996, p. 129). This pragmatism, it must be noted, is not utilitarian, because interests are conceived as local — it is not assumed that there exists a universal measure of utility. Also, the institutional aspects of scientific knowledge analysed by the sociologists of the Strong Programme, accounts for much of our intuitions that science is objective and truth oriented rather than utility oriented.

3.1.4 THE STRONG PROGRAMME'S ATTITUDE TOWARDS PSYCHOLOGY

Sociologists of the Strong Programme have multiplied the claims that their approach is compatible with psychological and cognitive approaches (see, e.g., Barnes 1976; Barnes et al. 1996, chapter 1; Bloor 1992; Barnes et al. 1996; Bloor 1997b, 1976, afterword of the second edition). They have recognised that these disciplines should have an important explanatory role in a science of science. As early as 1976, Barnes argued that the social sciences should not neglect the concept of natural rationality. In many places, Barnes and Bloor have argued that our collective cognitive accomplishments rest upon natural, individual inductive and deductive "inclinations" or "propensities". Members of the Strong Programme consider cognitive science and psychology as worthwhile empirical sciences, which can contribute to naturalising epistemology. Bloor dedicates a section of the afterword to the second edition of *Knowledge and Social Imagery* (pp. 165–170) where he denounces as meaningless the claim, attributed to him, that knowledge is purely or merely social. He then reassert his non-social reductivist position:

> The strong programme says that the social component is always present and always constitutive of knowledge. It does not say that it is the *only* component, of that it is the component that must necessarily be located as the trigger of any and every change: it can be a background condition.

Regarding the work on scientific heuristics in AI, he comments:

> The only sociologist to be upset by it would be those foolish enough to deny the need for a background theory about cognitive processes. I take it as evident that you could have no social structures without neural structures. Cognitive science of the type described is a study of just that background 'natural rationality' that advocates of the strong programmes take for granted.

More generally, the Strong Programme's understanding of the role of cognitive science is that:

> Cognitive science and the sociology of knowledge are really on the same side. They are both naturalistic and their approaches are complementary.

In spite of these claims, I know no study, within the Strong Programme's tradition, investigating the causal role of cognition in actual cases. The pscychological and cognitive conditions that makes the social processes of scientific knowledge production possible are taken for granted, but not investigated. In spite of explicit appeal to psychological phenomena, psychology itself is little exploited (but see, e.g., Bloor 1997b, were the work of Barlett is used); it mainly remains in the background. I think that this omission goes against the spirit of the Strong Programme and campaign for an integrated strong programme: a programme that aims at analysing all the explanatory causes of knowledge production, social, psychological or environmental. What are the reasons for omitting cognitive studies of science while their importance is acknowledged? The goal of the protagonists of the Strong Programme was first and foremost to show that the processes of scientific belief formation always included social phenomena. A prerequisite for achieving this goal was to defeat, or show the insufficiencies of, the competing methods in the history of science — traditional history of ideas and rational reconstruction. Yet, once this goal is achieved, one can go forward and study the cognitive bases of the social processes that issue and constitute scientific knowledge. For this project to happen, one must deal with more contemporary opponents: sociologists who are actually social reductionists, or who deny that psychology is relevant to science studies. Another impeachment of the integrative project is the fact that the violence of the reactions against the Strong Programme has deprived it from the help of psychologists, who have been prejudiced against it. Those who did cognitive studies of science, often thought important to demarcate their projects and views from those they mistakenly attributed to the Strong Programme. The mistake is understandable because sociologists of science actually disagree with this share of the tasks between sociology and psychology, and do indeed develop social reductionist theories. More radically, they even reject psychological explanations at the methodological level.

3.2 Alternatives to the Strong Programme

The Strong Programme proposes a method for studying science that first presented itself as a naturalistic alternative to rational reconstruction and teleological history of science. In the 30 years that separate us to the first spelling out of the Strong Programme (Bloor, 1976), social studies of science have provided many examples of the causal effect of conventions, institutions and other social phenomena on the production of scientific knowledge. Today, however, there exist numerous approaches, in social studies of science, which also claim to adopt a naturalistic perspective, but stand in opposition to the Strong Programme on several points. The existence of alternative approaches in social studies of science evidences the reasons why proponents of the strong programme should stop confining themselves to purely social studies and actually study the cognitive aspects of scientific knowledge production: integrating cognitive studies to the Strong Programme would allow exploiting the potentiality of its theoretical framework, thus showing its advantage over concurrent frameworks. Indeed, I will argue that the concurrent frameworks are ill-suited for the project of integration.

3.2.1 Rational reconstruction

Among the concurrent framework, there is, still, rational reconstruction. The Strong Programme has extensively argued that its own research methods and theoretical assumptions are better. Yet, the fear that the causal role of the non-social world and mental processes are being fully ignored has provided most material for the anti-Strong Progamme discourses. Beyond the anti-naturalistic reactions, one can often see a justified request that the cognitive aspect of scientific knowledge production be taken into account (e.g. Slezak, 1989). This craving for cognitive explanation stems from the blatancy of the fact that doing science requires much thinking, or (more generally) cognitive processing. Fortunately, the integration of cognitive explanations is possible while keeping all of the spirit and *strength* of the strong programme. This is even more so because the cognitive turn renders possible real naturalistic investigations of the psychological and informational processes at work in the practice of science – allowing, in particular, causal cognitive explanations without appealing to the normative aspects of rationality. If the discomfort science studies still generates can be snuffed out, I think it is by actually using cognitive science for pursuing and deepening their naturalistic enquiries. The problem lies in a rhetorical impossibility to convincingly argue that cognition has its own place in a Strong Programme analysis of knowledge production without clearly showing where this place is. The solution is not to give way to rational reconstruction and teleological explanations as a compromise to the critics, but to pursue naturalistic investigations on what the critics presume to be their own ground: cognition. We could have – and should aim at – a *unified*, *symmetrical* (both true and false beliefs have causal explanations) and *causal* account of scientific knowledge, which would consider both cognitive and social facts. What we would thus get is not some kind of compromise research programme. We would get a doubly strong programme – a programme that enlarges its scope and strength.

Also, the Strong Programme needs to provide arguments that show why its own methodological principles are really different from, and better than the alternatives within the Sociology of Scientific Knowledge. The main point of disagreement, I maintain, concerns the role that is given to cognition and, in particular, to the mental apparatus processing the input provided by the phenomena that the scientists investigate. Among the prominent trends in social studies of science, are:

Bath School whose main protagonist is H. M. Collins; it advocates methodological idealism

ANT whose main protagonist is B. Latour (but also M. Callon); it advocates the second symmetry principle

Ethnomethodology whose main protagonist, in science studies, is M. Lynch; it advocates *describing* practices rather than seeking causal accounts

The practice turn characterise a movement in science studies, which emphasis the analysis of scientific practices, sometimes at the expense, I would argue, of the analysis of beliefs. Pickering's edited book *Science as Practice and Culture* (1992) is a milestone of this turn.

social epistemology it is a trend in normative epistemology, which takes into account the description of the social processes in science in order to prescribe social actions. The most developed school of that trend is feminist epistemology.

I now review briefly these alternatives, but I will not consider the trends in social epistemology, are their goals differ from the descriptive and explanatory goal of an integrated science of science.

3.2.2 The Bath School

The Edinburgh school rejects Collins' method (1992), which consists in a methodological denial of the existence of an outside world impinging on us and being described by scientists. According to Collins, "all descriptive-type language should be treated as though it did not describe anything real", but as though it was about imaginary objects. Barnes, Bloor, and Henry (1996) applauds the *goal* of the method, which is to uncover assumptions that are taken for granted, but denounces the method itself for being at odd with the fact of the matter: the outside world does impinge on the scientists' senses and hence has a causal impact on the scientists' descriptions of it. The problem with Collins' principle is that it prevents from taking into account the causal action of the phenomena investigated on the beliefs of the scientists. By denouncing Collins' principle, therefore, Barnes, Bloor and Henry reaffirm that the causal relations between the world and the scientists are, in the perspective of the Strong Programme, worth investigating. Collins' principle is motivated because describing the causal effect of the world on scientists' belief formation requires some description of the world, such as "Newton saw the apple falling". So Collins rightly fears that this description will lead the analyst to already assume what is to be explained, as for instance in "Newton saw the apple being attracted by the earth". Contra Collins, the way out is not to ban any description of the world since it has the damaging consequence of unduly discarding some causal factors. The way out consist in choosing carefully what can be assumed and used in the explanation and what must be explained. In truth, the necessity to assume some propositions in one's scientific investigation on knowledge production is the fate of naturalistic epistemology and the main point of relativism – but no-one, by now, should be ashamed to work on Neurath's boat [1]. The Strong Programme's rejection of Collins' principle directs us towards cognitive studies of science: The analysis of the causal role of the world on knowledge production requires going back to what the scientists have sensed and to the cognitive processing processing of the input provided by the phenomena investigated. This means doing cognitive psychology. Barnes et al. (1996, p. 75–77) review Collins' contribution to understanding how the problem of induction is solved, in context, by scientists: regularities are perceived through the application of socially established conception of regularity in describing nature. Barnes et al. recognise the importance of prior knowledge and social established conception in interpreting regularities, but disagree with Collins' offering his account as a *replacement* of the individualistic accounts of the psychologists.

> *Pace* Collins and his interesting idealist arguments, sociologist should be willing to acknowledge the existence and the causal relevance of the physical environment when they study the growth of knowledge. And having acknowledged this, they should acknowledge also the ability of individual human beings to monitor the physical environment and learn about it.

[1] The metaphor compares someone wanting to keep his boat afloat while examining the planks of the hull. The idea I drawn upon is that, in this situation, one can never examine the whole of the hull at the same time, but needs to look at the planks one by one while keeping the other planks in place so that the boat keeps floating. Likewise with epistemology: one cannot question the whole of knowledge at once, and needs assume most of it in order to study. This is a consequence of the naturalistic, non-foundational, programme.

> Individual animals learn directly from experience. The psychologist's rat pushes the lever and looks of the arrival of a food pellet [. . .] The rat has learned to associate the lever movement with the arrival of food, and has developed an expectation that this association will continue in the future. The rat has successfully operated as an inductive learning machine: perhaps we should credit it with 'inductive reasoning'. In any event, it would be perverse to insist that what rats manage to accomplish in this context, human beings cannot hope to emulate [. . .] It is more plausible to accept that the human brain, like the rat brain, may be profoundly affected by the reception of signals from the physical environment, and that these signals may directly engender not just perceptions and memories but associations and expectations as well.

3.2.3 Latour and ANT

The Actor Network Theory (ANT) raises in some way Collins' methodological principle to an ontological commitment. In ANT's perspective, the world itself, rather than just our knowledge of it, is socially constructed. At the basis of this move is the will to question, or simply deny, the subject-object distinction and the notion of representation as being about something (Woolgar, 1989; Latour, 1987). Thus, Latour (1987, p. 258) provides the following "rule of method":

> Since the settlement of a controversy is the *cause* of Nature's representation, not its consequence, we can never use this consequence, Nature, to explain how and why a controversy has been settled.

Note that "Nature's representation" used in the first part of rule 3, is transformed in "Nature" in the second part. They are assumed to be the same thing!

The Sociology of Scientific Knowledge has often been accused of forgetting the role of the world on the formation of scientific knowledge

(Bricmont and Sokal, 2001). The solution of Latour and Woolgar is to conflate the world with our representation of it (Latour, 1987; Woolgar, 1989). Their consequent methodological advice is to forbid treating the phenomena investigated by scientists as input from the world determining the cognitive processes of some cognitive system – a scientist's mind or a laboratory – that produces scientific representations. In the framework of Actor Network Theory, all are 'actants' in an undifferentiated network. Undifferentiating types of 'actants', first, hinders calling on specific theoretical resources, such as cognitive psychology for understanding the mind's processes, and second, blurs the distinction, essential for explanation, between explanandum and expanandans. For instance, microbes are not part of the cognitive system that allowed 'the pasteurization of France' as would suggest Latour's account (1993), but they provided the input. In the traditional view of the Strong Programme, the explanandum is the beliefs and practices of scientists, rather than microbes themselves, and the explanans includes the social context, and it also includes other types of causes to be found in the constraints and the constructive role of the mind when it processes the stimuli provided by the world. So the phenomena under scientific investigation have a causal role when they stimulate our senses, or, more generally (since scientific artefacts often mediate between the phenomena and the human senses), when they provide an input to the distributed cognitive systems of science. The main difficulty for the scientist of science is to avoid describing the input in the same terms that are used for the output.

The opposition between ANT and the Strong Programme is therefore more radical, since it bears also on what the sociologist of science needs to explain. In the perspective of the Strong Programme, the sociologist of science is to discover the causes of scientists' beliefs about the outside world. It thus fully relies on the subject-object distinction and, moreover, set cognitive phenomena — beliefs — as its proper object of investigation. Of course, the social distribution of beliefs is of outermost importance for the sociologist of the Strong Programme, e.g. it enters their sociological definition of. But it remains that beliefs are cognitive mental objects. The Strong

Programme thus fully acknowledges that exhaustive explanations, in science studies, need to rely on cognitive science. This contrasts strongly with Latour and Woolgar's call for a moratorium of cognitive studies of science (Latour and Woolgar 1986, p.280; Latour, 1987, p.247), which boldly asserted that science could be better explained without cognitive science.

Latour's second symmetry principle prescribes suspending our belief in a distinction between natural and social actors; it goes against the dichotomies between object and subject, and between nature and society. The second symmetry principle is presented as an answer to the unappealing social reductionism, which is (wrongly) attributed to the Strong Programme. He says:

> To be sure, Bloor is not an idealist—as are some of the other descendants of the Edinburgh school—and for him the social is only one half of the explanation, but the other half is completely unclear. I think now the only way to achieve Bloor's goal is through what Michel Callon calls the generalized principle of symmetry.

Maybe Latour is right to complain that one of the explanation is unclear; but the clarification is to be found in empirical, cognitive psychology, rather than in a principle with un-intuitive metaphysical commitments.

3.2.4 Ethnomethodology

Ethnomethodology has also presented itself as an alternative to the Strong Programme. Thus Michael Lynch Lynch (1992) asserts that "we have good reason to give up on causal SSK and turn to ethnomethodology". There are two points that ethnomethodologists have taken as distinctive of their approach. The first is their commitment to study the details of "situated practices in science" or "scientific practices in-their-course", which they assert to be highly relevant to science studies; the second is their condemnation of causal explanation and their vow to confine themselves to description. With regard to the first point, one can easily observe that ethnomethodologists do not have the monopoly for the study of scientific *practices*. The sociology of Edinburgh, in particular, pays special attention to practices as both determinants of scientific results and situated in specific traditions. However, ethnomethodologists may distinguish themselves with their investigation of the "embodied, communicative performance of social and natural scientific methods" (Lynch, 2001) and the mundane taken-for-granted practices. Could one incorporate the ensuing detailed descriptions in a causal account *à la* Strong Programme? Yes, provided that we move into cognitive science: recent work on embodied, situated and distributed cognition and cognition in action allows analysing the situated practices together with their mental underlying causes. In this perspective, the cognitive human agents think not only about the environment, but also *with* the environment, thus grasping the importance of practices as situated actions. Moreover, the practices are analysed not only as "producing social orders", but also, and essentially in the sciences, as producing representations of the phenomena investigated by the scientists. Then, if the study of situated practices is not sufficient to distinguish ethnomethodology, then the most distinctive point is their self proclaimed ban on causal explanations. This ban does not come from a supposed flaw with causal explanations *per se*, but stands on a philosophical argument which asserts that the phenomena investigated by ethnomethodologists, rule applications (or the implementation of methods), cannot be grasped by causal explanation because there are no such causes as the ones postulated by the Strong Programme (Lynch, 1992). The renewed debate on this topic, initiated by Kusch (2004), is revealing: Bloor (2004a), once again, needs to point out that a causal reductive account of rule following cannot remain at the level of social institutions, but must eventually account of the latter with the psychological processes that generate and sustain them, through individual's actions and together with social interactions. Bloor's counter-argument to the ethnomethodologists therefore consists in an appeal to the causal explanations that psychology can provide.

3.2.5 The practice turn

The practice turn is often conceived as going beyond the sociology of the Strong Programme, which would focus on science-as-knowledge rather than science-as-practice (Pickering, 1992), and provide an erroneous image of science. ANT and Ethnomethodology would have well taken the practice turn, while the Strong Programme remained behind. There is much to take from the practice turn and its new emphasis on practices and the situatedness of scientific actions, the role of know-how, and the importance of technology. I will dedicate the third part of this thesis to these important aspects of scientific knowledge production. However, one could regret that the study of practices is often done at the expense of the study of the processes through which scientific beliefs are formed. In fact, beliefs do determine practices to a great extent. The initial focus of the Strong Programme on belief formation is therefore far from misguiding research. It is so only in when beliefs themselves are thought as entities that are not worth studying—which has probably been the idea of those who abandoned the study of science-as-knowledge for the study of science-as-practice.

3.2.6 Distinguishing the schools of social studies of science

For those who agree that a central project of science studies is to provide causal explanations of why scientists believe what they believe, the Strong Programme advocates a better methodology than the above mentioned alternatives in science studies: rational analysis, Collins' methodological denial, Latourians principles and ethnomethodological descriptions. The strong Programme provides a better methodology because it allows integrating the concepts and results of those who have dedicated their research to the analysis of the mental phenomena causally involved in belief formation. However, it is only by showing the fruits of what is proper to the Strong Programme - its integrative potential - that what distinguishes it from Collins, Latour and Lynch's programmes can be made more than a question of abstract principles, and the theoretical disputes more than philosophical quibbles. In other words, members of the school of Edinburgh should actively work for the integration of cognitive and social studies of science and put hands on cognitive studies. With the integration of cognitive studies of science only, will the Strong Programme harvest more research results, more comprehensive and satisfactory explanations of particular cases, than the alternative theoretical framework in social studies of science.

More often than not, opponents to post-Mertonian developments in the sociology of science do not distinguish the different theories and principles of the different schools in social studies of science. They would build their some straw man by picking the most outrageous claims, and conclude that the Strong Programme is definitely mistaken. Let me quote Hull (1988) as an example:

> Thus, Collins's (1981a: 218) assertion that sociologists of science, in their investigations, "must treat the natural world as thought it in no way constrains what is believed to be" can serve as a useful antidote to our usual prejudices. But how about sociologists of science themselves? Advocates of the strong programme urge extensive empirical investigations of the actual practice of science. But to what end if the natural world in no way constrains our beliefs? — Hull (1988, p. 4)

In this quote, Hull makes two mistakes mentioned above: the methodological principle is first recognised as a useful antidote — which is what it is meant to be, but Collins' '*as though*' is forgotten two lines below when Hull seems to attribute the belief that the natural world 'in no way constrains our beliefs' to the strong programme. But doing as though the natural world does not constrain the behaviour of the scientists that is being studied does not imply believing that it is so, and the sociologist of science can with no qualms study scientists' behaviour. The second mistake is that the methodological principle is formulated by Collins who does not belong to the strong programme, often called the Edinburgh school, but to the Bath school. We have see that Collins' methodological principle has been strongly criticised by the Edinburgh school as misleading, because it unnecessarily prohibit the study of the actual causal action of the natural world in belief formation.

I admit that diving into the details of the schools in SSK and their particular methodological principles may not be appealing for someone who already feel repealed by the first account he has had of what is being done. Most cognitively oriented readers may recognise themselves on that point. This is where I might be useful: I have been doing some of the work for psychologists of science; I have done a critical analysis of the schools in SSK and put into evidence what I think are the most fruitful and psychologically friendly claims in SSK.

In this section, I have argued that integration was *the* research strategy that could allow specifying and clarifying the distinctive principles of the Strong Programme and their fruitfulness. In the next section, I will argue that integration would provide empirical grounds to the principles of the Strong Programme.

3.3 NATURALISING EPISTEMOLOGY: VIA SOCIOLOGY, TOWARDS PSYCHOLOGY

The main and central goal of the Strong Programme is the naturalisation of epistemology, which is understood as the requirement to provide explanations of scientific knowledge with the natural causes that produce it. The second goal of the Strong Programme has been to show that a natural account of scientific knowledge production would inevitably and constantly rely on sociological inquiries. The Strong Programme can thus be presented as having two components. The first is a naturalistic stance for the study of scientists' claims and actions. It is a methodological point. The second assert that social phenomena always have a determinant role in scientific knowledge production. It is a theoretical point. A corollary of it, however, is that psychological facts cannot, on their own, fully determinate the production and content of scientific knowledge. In the following I will attempt to disentangle some of the social and cognitive factors in scientific knowledge production.

3.3.1 CANDIDATE SOLUTIONS TO THE UNDER-DETERMINATION PROBLEM

How can social phenomena have such a pervasive role as the one claimed by the Strong Programme? A standard answer given by social scientists is to assume that the mind is totally framed by the social, through enculturation. From that point of view, 'the social' is given the best role, while psychological facts are relegated to providing the material basis (the brain) for the otherwise social determinations. Proponents of the strong programme, however, are not committed to such a view. They have shown their belief that the mind include processes that are independent from, or unaffected by, social enculturation when, for instance, siding with Fodor's modular theory of perception against Churchland's view (Barnes et al., 1996, chap.1). The Strong Programme's argument showing the pervasiveness of social events in scientific knowledge production does not depend on the assumption that the human mind is wholly shaped by culture.

The strategy of the school of Edinburgh's for showing the all-pervasive causal role of social interactions in scientific knowledge production has been to invoke the underdetermination thesis. In its primary form, this thesis states that scientific beliefs cannot be derived logically from data only, that there is no logical procedure that allows choosing among competing theories and that there always are auxiliary theories between any theory and the data (theory ladenness). These points first show the explanatory limits of rational reconstruction and the necessity to find out the causal determinations of scientific practices for explaining science. Underdetermination, however, does not imply that the missing determinants should come from social factors. In fact, the underdetermination of agents' response by the stimuli has constituted an important argument for cognitive scientists who have posited the existence of rich innate cognitive structures as providing the missing determining constraints. The now archetypical example is Chomsky's universal grammar: a human innate capacity without which children could not learn the grammar of their language; the latter being undetermined by any set of utterances a child can hear. In this case, as in many cases of learning, the underdetermination problem is wholly solved by biologically given cognitive constraints, thus letting little role to social factors.

Scientific practices might be wholly determined by non-social factors: the mere result of the human mind directly — i.e. without social biases or socially framed presuppositions — interacting with the outside world. Of course, social interactions enter in a trivial way in scientific knowledge production since science is not the product of a single mind but has evolved in history, being passed, so to speak, from one scientist to another. But the Strong Programme wants to show that social interactions have further role than just allowing some ratchet effect. According to it, scientific knowledge inherited from the past is not only a set of results that are to be taken into account in further scientific development, it also constitutes a cultural tradition — possibly interacting with some other cultural traditions and social interests — which informs future applications and development of knowledge. Accordingly, each case of scientific knowledge production causally involves the social context — be it scientific (local) traditions or other social factors stemming from the culture in which the scientist is immersed.

The most elaborate illustration of the above assertion has been done on word application (see, e.g. Barnes et al., 1996, chap. 3): meaning finitism points out that every application of a word is justified by the identity of the thing or the event named with the previous things or events that have been named by the same word. But everything is always different in some ways to other things, so there is again this logical possibility to question any application of a word. Here, cognitive determinations are the obvious candidate as causal factors for word application: our feelings of identity, our intuitions will determine our use of words. For instance, if this liquid looks like water, smells like water, tastes like water, then, we will call it 'water'. Psychological constraints that make the learning of words possible are indeed necessary and attempts to identify them have led, for instance, to posit the existence of a Language of Thought or a set of assumptions that allows children to pick the reference of names among the numerous logical possibilities opened by ostentation. As it happens, however, our feelings of identity are not sufficient on their own to determine word application *in science*: intuitions and feelings of identity can always be questioned. For instance we may realise, but only after some chemical test, that some liquid is not H2O but, say, XYZ. It is not that intuitions and feelings of identity have changed; they have remained the same, but have been made irrelevant for word application by the newly created scientific context. Some other intuitions are applied for discriminating water from XYZ. For instance, if XYZ turns red while water turns green when adding some other substance, then the final judgement will rely on colour perception (plus the idea of having rightly manipulated the substances). The beliefs about what counts as evidence and what does not, the acquired skills in bringing about evidence, the argumentative techniques for persuading colleague scientists are all cognitive effects of social interactions that are causally involved in the scientific judgement of word application. So, while the judgement of an individual scientist is still determined by his thoughts, a causal history of the origin of the thoughts must extend outside of the brain of the scientist and take social factors into account. Social interactions inform scientists of the decisions procedures: taking some intuitions or percept as non-discriminative and some others as furnishing, in a given context, the criteria for word application. The scientist interprets and uses his own intuitions through local interpretive traditions which are, at least in science, the product of social interactions.

Without postulating an exaggerated plasticity of the mind, sociologists of Edinburgh still account for the pervasive role of social factors in scientific cognition. This is because the basic intuitions that are caused by the innate endowment of the human mind are interpreted in the light of the historical social context. Scientific responses to stimuli are therefore underdetermined by facts about the mind. The underdetermination argument thus understood also hits individualistic psychological explanation. So we have two kinds of essential and pervasive causes of knowledge production: causes originating from facts about the human mind and causes stemming from social interactions (a third cause, of course, being the nature of the stimuli provided by the phenomena observed). The Strong Programme, however, has not yet spelled out how to combine theories of the innate aspects of cognition with the above social constructivists ideas (I qualify this component of the Strong Programme as social constructivist, although the label is unevenly used by the sociologists of the Strong Programme). It has not provided an account of the relative role of sensory stimulation and cognition, how it causally intervenes in knowledge production and how it relates to social factors. Barnes, Bloor and Henry devote the first chapter of *Scientific Knowledge: A Sociological Analysis* to a debate in the cognitive science between Churchland and Fodor regarding the cognitive processes leading to perception. Fodor argues that these processes are innately specified and cannot be changed or informed by our beliefs. Churchland, on the contrary, holds that what we perceive change with the progress of knowledge. Barnes, Bloor and Henry eventually side with Fodor's modular theory of perception. They then clarify at which cognitive level social factors step in: since perceptive cognitive processes are unaffected by one's beliefs, it is at an upper cognitive level that culturally acquired beliefs step in the individual scientists' production of scientific thoughts. So the second chapter is devoted to the role of culturally acquired beliefs at this upper cognitive level, which they call 'interpretation'. Unfortunately, the cognitive processes mentioned in the first part are not given attention anymore, and the analysis considers only the content and structure of scientific arguments, pointing out the necessary presence of the "local interpretive tradition" without showing its role in actual cognitive processes. The consequence is that while the Strong Programme has repeatedly stated that sensory stimulation by the non-social world and cognition play a central role in knowledge production, it has not yet convinced its detractors. There are two different levels of underdetermination: at the first level, stimuli, sensory stimulation, underdetermines our thoughts about the world. This first underdetermination allows refuting the foundational project of logical empiricism and requires investigating the provenance of complementary determination. The causal factors sought are presumably psychological, sociological or both. Works in cognitive psychology show that there are psychological determinations, and works in sociology of scientific knowledge show the social determination. The right answer is therefore that both psychological and social factors act on the formation of scientific beliefs [2]. The task remains to show how social and psychological factors relate to each other, what the role of mental capacities is, and what the cognitive processes that relate their output with the interpretive traditions are. This is what I will do in the next part of the thesis.

3.3.2 Methodological individualism as a naturalistic method for epistemology

A requirement, for plugging in psychological studies onto social studies, is that the individuals' behaviour be revealed in the account of social phenomena. Holistic or macro-social approaches in the social sciences do not satisfy this requirement. Where does the Strong Programme stand on this point? It seems that the Strong Programme's assertion about the irreducibility of social phenomena in knowledge production is often confused with the assertion that social phenomena are themselves irreducible [3]. Much of the work of Barnes and Bloor, however, bears on the understanding of social phenomena as emerging from individual's behaviour; it provides an analysis of the individual's action that create the institutional character of science.

The all pervasive presence of social interactions and cultural situatedness in scientific knowledge production has led the sociologists of Edinburgh to analyse scientific knowledge as one form of institution. This al-

[2] Some philosophers, such as Giere (1992) or Kitcher (1998), have accused sociologists of science of overinterpreting the Duhem-Quine thesis. It is true that the thesis has been used indistinguishably against both foundationalist theories such as logical positivism and against more current individualistic epistemology. I hope that my distinction between two levels of underdetermination clarifies the argument for sociologism.

[3] Kitcher (2000), for instance, mocks Bloor's appeal to Durkheim. He seems to believes that the Strong Programme aims at finding 'social facts' as irreducible entities, while it aims at providing sociological explanations. As a matter of fact, Durkheim's work may be of interest even for methodological individualists (see, e.g. the work of R. Boudon).

lows them to account, first, for the normative aspects of science as a requirement to comply with the conventions of the knowledge institutions and, second, for the processes of social coordination leading to the conventional aspects of knowledge making. At this stage, the sociological stance of the school of Edinburgh is radical, Bloor not hesitating to talk, in the manner of Durkheim, of 'social facts'. However, this radical move towards sociology calls, once again, for further investigation of its relation with cognitive psychology. This is because the naturalistic stance of the Strong Programme involves clarifying the ontological status of the postulated entities. A naturalised sociology, indeed, aims at showing that institutions or social facts *can* be understood in terms of entities that are not themselves institutions or social facts. Such entities are people, their behaviour and their environment, which furnishes the empirical ground to sociological analysis. Barnes (1983) and Bloor (1997a) have developed a model of social institutions that is reductive in the sense that it accounts for institutions in terms of interacting individuals. The simplified model requires only that the agents be able to coordinate with each others while having their own judgemental or discriminative abilities. The coordination implies that the agents intend and succeed to do as the others do or expect them to do (they are social agents). This understanding of social phenomena therefore goes back to individuals' action – It involves micro-sociology. This is well illustrated in the case studies of the Strong Programme, where changes in science, such as a change in paradigm, are explained with the scientists' thoughts and interests. These explanations can be distinguished from macro-analysis, whose explanation rely on institutional change only (e.g. policy of some University) or technical change. The sociology of Edimburgh takes on the task to investigate the individuals' actions that form the basis of social phenomena.

Barnes and Bloor's model of institution and account of social phenomena belong to methodological individualism, as opposed to holist theories in the social sciences. They analyse social phenomena in terms of properties emerging from the actions and interactions of individuals. Their methodological individualism, however, does not rely on traditional rational choice theory and its associated fully rational economic agent, as is often the case in methodological individualist sociology. Traditional Rational Choice Theory endows the agents with a rationality that is the exact replicate of current scientific norms of reasoning: logical consistency, calculation of probability using the latest development in Bayesian theory, etc. It is therefore not surprising that Rational Choice Theory so conceived is not used as a theoretical resource by the sociology of Edinburgh, which proscribes the explanation of behaviour with a normative rationality turned as a descriptive tool for the sociologist's purpose. This would amount indeed to the kind of Rational reconstruction argued against and to explain the generation of rational norms by putting them in the agent beforehand. But if the traditional model of rational agent of Rational Choice theory cannot enter Edinburgh's methodological individualist account, on what resources shall it draw for the analysis of scientists' behaviour? The obvious answer is to turn towards psychology. It is to integrate the Strong Programme. From our discussion on under-determination, we know that social events and public representations are necessarily taken as input of the mental processes that generate individuals' scientific beliefs. Now, the mental processes themselves are (tautologically) to be studied by psychology. Also, the extent with which social inputs impinge on the processes themselves is an empirical question that pertains to psychology. So we have a line of investigation whose methodological consequences lead to the theoretical resources of psychology.

With the goal of naturalising epistemology , the school of Edinburgh has shown the important part played by social interactions in scientific knowledge. It has argued that one cannot naturalise epistemology without sociological enquiries. Yet, while it is compulsory to go first through sociology, the naturalising process does not end there – it requires continuing towards psychology. This is because the Strong Programme's 'sociologism' eventually leads to the assertion that scientific beliefs of individual scientists are the outcome of psychological processes that importantly include the scientists' cultural background in their input. I maintain that the cultural determination of scientific cognition can be understood only by taking into account the way the mind processes both social and non-social inputs. Cognitive determinations happen both in the scientists' production of public representation and in the assenting or dissenting behaviours of the audience. Also, the very institutional character of science relies on cognitive foundations: people are able to communicate with each other and to coordinate their actions. They are able to agree on scientific problems and continue to use the given solution in an unproblematic way during long periods of time. The principles of the Strong Programme, however, still need to be specified and fleshed out in their psychological assumptions. The specification would open the way to integrated studies of science, taking advantage of the explanatory power of both cognitive and social sciences. The ensuing explanations would, without doubt, show the superiority of the methodological principles of the Strong Programme over its competitors in social studies of science.

My methodological proposal is a doubly radical programme: radical because it strongly takes on the naturalistic anti-dualist and reductionnist view of the mind developed by cognitive science; and radical because of its attempt to develop naturalistic alternatives to rational reconstruction for understanding scientific developments. For more than a generation, studies adopting the Strong Programme have inquired about the social facture of science in a thorough and persistent way. This led Latour (1999b) to mock Bloor for "not moving an inch". To some, it may therefore be strange to still call on the strong programme for developing new methodological principles for science studies. Yet, insofar as the programme consists in the uncompromising scientific inquiry of scientific beliefs, then no metaphysical turn is necessary or desirable to bypass the limits of the sociology of scientific knowledge. What is desirable, however, is that all the explanatory power of scientific disciplines be put at work when relevant. Starting from the sociological perspective of the Strong Programme, I have attempted to show the relevance of cognitive science to science studies.

Apart from being a field of studies of its own, cognitive studies of science can clarify, explicate and fruitfully challenge concepts and theories used in social studies.

Part II

Psychology and the history of science

Cognitive history of science attempts to describe the cognitive bases of scientific evolution. It asks, in particular, what cognitive abilities are put to work in scientific thinking, and how these abilities enable and constraint the evolution of science. Sperber, Hirschfeld, Atran, Boyer and others has shown how cultural stability and variability can be based on species-specific cognitive capacities. Their explanations of cultural phenomena are informed by theories in developmental and evolutionary psychology, which have described the human mind as endowed with domain specific evolved cognitive abilities. The four chapters of this part are at the crossroads of cognitive history and the above work in cognitive anthropology. The evolution of science is an instance of cultural variation; it is an instance of evolution in the distribution of representations among the population. How are scientific representations produced and distributed? And what is the role of the mental apparatus in the processes of production and distribution?

My main argument is that human evolved cognitive abilities enable scientific evolution because they allow interpreting new data in the light of some culturally acquired framework: this implies that social factors determine the content of science. The processes of scientific belief formation result in variable and evolving beliefs in part because they depend on the context. But in spite of this potential variability, psychological factors also determine the content of science. I present the epidemiology of scientific representations as a theoretical means to specify the psychological constraints on scientific knowledge production.

> Psychological considerations are crucial to a proper characterization of what is cultural. In particular, if one factors a natural-

> istic account of culture (as all evolutionary theorists and a few anthropologists do), then the naturalistic account of the mind that is currently developed in cognitive science should be of obvious relevance. Psychological considerations are also crucial to a proper explanation of cultural facts because psychological factors do more than enable culture, they contribute to shaping it. (Sperber, 2006a)

In this quote, Sperber evokes two reasons for bringing in the findings of psychology into social studies: it would enable (1) a proper characterisation and (2) a proper explanation of what is cultural. In the next chapter (chap. 4), I pursue the goal of providing a naturalistic characterisation of scientific evolution. I specifically deal with the problem of rationality, which is a notoriously hard notion to naturalise, but which is entangled (and rightly so) with understandings of scientific cognition.

The second reason why psychological considerations are important, say Sperber, is because psychological factors shape culture. How is it possible? And does that hold also for science? It seems uncontroversial that the nature of the mind be consequential on its production — science included. Yet, the advent of relativity theory and Riemannian geometry has showed that Kant had mistakenly attributed to Newtonian mechanics and Euclidean geometry a necessary status on the basis of their purported relation with the structure of the human mind. Since then, few have ventured to point any relation between scientific knowledge and psychological phenomena (notable exceptions, however, are still found in the philosophy of mathematics — see section 7.1). Couldn't we say that the evolution of science consists in bypassing the limits of our minds in order to grasp the real world? But if these limits are bypassed — as I think they are — then why should they be relevant to the study of science? The thesis that scientific knowledge is produced by the mind but independent in content from psychological factors has a certain appeal; while believing the contrary has been taxed of being psychologistic. It seems that if scientific cognition is the mere implementation of scientific methodologies, then psychological considerations cannot explain scientific knowledge. Suppose, for instance, that cognition is scientific only when it obeys the norms set by logical empiricists (e.g. some inductive or probabilistic method), then psychology is relevant only to the extent that in can explain how scientists manage to obey the norms (e.g. scientists are rational so they naturally obey the norms). With such a characterisation of scientific cognition, it seems that psychologists have nothing to say on the content of science: psychological factors enable science but do not shape it. The shape of science is accountable with the norms of logical empiricism and the data processed according to the norms. The independency thesis — which asserts that the content of science is independent from psychological factors — is a consequence of the idea that scientific thinking is the application of epistemic criteria that are independent of psychological factors. The thesis can be drawn whether these criteria are thought to come from a transcendent rationality or are social constructs.

The aim of this first part is to show that the independency thesis is false; and that it is false even when, or rather especially when, one adopts some kind of social constructivism. The argument showing that psychological factors shape the content of scientific knowledge proceeds in two stages: first I argue, in chapter 5, that there are psychological factors that are relatively independent from people's specific environment, especially their cultural environment; second, I show through an analysis of some mental mechanisms occurring in the distribution of scientific representation, how these factors can shape the content of science. The chapter 6 is a case study in the history of mathematics: I analyse how evolved mathematical skills may have influenced the course of mathematical theorisation of the calculus.

Chapter 4

Scientific thinking and rationality

In this chapter, I analyse the psychological grounding of the view defended in social studies of science, according to which scientific knowledge result from the local circumstances and social context of its production, rather than from the application of a-historic norms of scientificity. The main objective is to provide a psychologically viable account that does not indulge in *ad hoc* theorisation aimed at fitting some pre-established model of social change. The battleground of the argument is the debate over rationality: the notion of rationality has been used as a shovel to dig the gap between studies of scientific cognition and studies of the social aspects of scientific knowledge production (sect. 2.2). I will argue that for some actors of the rationality debate the opposition stems from referring to different things when talking about rationality, rather than making incompatible empirical claims. On this basis, I begin to sketch a view of scientific knowledge production, which asserts that scientists exploit much the same cognitive processes across the history of science, and that social processes importantly determine what counts as scientific. The view draws both on cognitive universalism and social constructivism.

The first section of the chapter is an essay in analytical philosophy about the notion of rationality; the second section attempts to fructify the analysis by combining claims from Sperber and from Barnes and Bloor; the third section specifies properties of the mind that are relevant for understanding scientific cognition, and insists on their compatibility with some trends in the sociology of scientific knowledge.

4.1 RATIONALITY: MIND AND CULTURE

The opposition between relativism and rationalism has hindered the integration of cognitive and social studies of science (see sect. 2.2). Sociologists of science are usually relativists, and those studying scientific cognition largely rationalists. Beyond these disputes, it is admitted that the scientific enterprise involves both scientists' brains and a great amount of social interactions. But this obvious fact is occulted when one gets to qualify scientific practice as a rational practice. As such, this qualification is tautological: scientific and mathematical practices have always been used as exemplars of rational behaviour. But because of the sacrosanct term 'rational', its multiple connotations and its associated disputes, the tautology unduly divides science studies.

Methodologies for the study of science are often derived from preliminary conceptual work on the notion of rationality. This is mostly how theorizations of knowledge production and acquisition have proceeded from Platon till Bloor or Kitcher. The analysis I am advancing consists in distinguishing several phenomena philosophers refer to when talking about rationality, then setting up the scientific task to find out how these phenomena are related. This section includes a shopping session of theoretical positions with names ending in 'ism'. But I try to clearly define the positions and justify my choices. The final goal is to specify a research programme for describing the rational character of scientific cognition. The resulting research programme consists in specifying the path from the biology of the brain to social norms of good thinking, from the human cognitive apparatus to behaviour to culture and historical developments, and conversely from norms of good thinking to the biological apparatus that enable us to comply with the norms.

4.1.1 The psychologist, the epistemologist and the social scientist about rationality

When one argues against empiricism, 'rationality' is understood as designating the cognitive capacities with which normal humans are endowed. One insists on the fact that the mind has its own material with which to think, and opposes this view to the idea that all that is in the mind come from the senses. When one argues against relativism, a second meaning of 'rationality' is often used: 'rationality' is understood as designating the principles of *good* thinking. Is rational what is well thought. Against the relativist, the rationalist argues that there exist ways to attain true and justified beliefs, and that these ways are independent of the person who thinks or his culture. With these two meanings, rationalism can either be understood as a psychological thesis about the nature of the human cognitive apparatus, or as an epistemological thesis that asserts that there is one predetermined context independent set of ways to reason well. In the first case, the rationalist is the enemy of empiricists; in the second it is the enemy of relativists. In some debates, these different meanings of 'rationality' and 'rationalism' are undistinguished, leading to some deadlocks, as when rationalist philosophers working on the psychology of science and relativist sociologists of science talk past each others.

It is therefore important to distinguish two questions: one about the plurality or the uniqueness of actual psychological processes; the other about the plurality or uniqueness of ways to think well. Stich (1990, p.13, 14) names the alternative consequent views as follow:

descriptive cognitive pluralism "different people go about the business of cognition – the forming and revising of beliefs and other cognitive states – in significantly different ways"

descriptive cognitive monism "the idea that all people exploit much the same cognitive processes"

normative cognitive pluralism "there is no unique system of cognitive processes that people should use, because various systems of cognitive processes that are very different from each other may all be equally good"

normative cognitive monism "all normatively sanctioned systems of cognitive processing are minor variation of one another"

The two first positions are psychological; they are empirical assertions about the human mind. For instance, descriptive cognitive pluralism implies that the human mind is sufficiently plastic so as to take different structures when exposed to different stimuli. In particular, the cultural environment is, according to some relativist anthropologists, thought to generate different ways to think. The two last positions – normative cognitive pluralism and monism – are epistemological; they are claims about good cognitive processes, the processes people ought to use. Monism and pluralism are matters of degree. They state on the similarity between different cognitive processes and different norms of good reasoning; these processes and norms are more or less similar.

The epistemologist has to choose between normative cognitive monism and normative cognitive pluralism; and he may want to decide in function of which is true between descriptive cognitive monism and pluralism. For instance, normative pluralism has had much support from descriptive pluralism when post-colonialist anthropology attacked the ideas that indigenous thinking was not passing the standard of good thinking and qualified as 'primitive' or 'pre-logical'. There are therefore the descriptive and the normative levels, and the descriptive level is of consequence to the normative level, especially for works in naturalised epistemology (e.g. Goldman, 1986).

There is yet another level, which is made of the actual epistemologies held by the people, i.e. what they think is good thinking. This level can be described. Explicit epistemologies are described by historians of philosophy, who analyse the thoughts of Descartes, Hume, Carnap, Popper, etc. about the ways to attain knowledge. The historians of science also describe the epistemologies held by the scientists. This is important because scientists' beliefs about what is good reasoning influence scientific production, and because scientific production always include prescriptions about how to think well about some subject matters. The history of the judgements about what is rational or scientific is at the heart of science studies.

What are the empirical data that sociologists of science gather when investigating the norms of good reasoning used by the scientists? Scientists' ideas about what good thinking is is revealed in their approval or disapproval of thought processes. These are social phenomena that are observable facts that enable social scientists to circumscribe what is taken to be rational thinking at a given time in a given community. In this view, ideas about rationality have a normative effect in the sense that they implement social norms; whether a thought process is understood as rational or not depends on whether the community in which the thinking took place approves it as valid or not. While rational behaviour obviously implies cognitive processes, the adjective 'rational' is attributed to those cognitive processes whose output are deemed to conform to a social norm. This is why there can be errors: not all outputs conform to the norm. So, while the epistemologist is interested about ways to reasons well, the scientist of science is, by contrast, interested about *ideas* about ways to reason well, and how these ideas determine knowledge production. The goal of the scientist of science is to describe what is taken to be good thinking; the description takes the form of interpretations of scientists' epistemological ideas and thoughts about what is rational, which can be qualified as *tacit epistemologies or rationalities*. Importantly, these thoughts are included in most knowledge claims.

The question that consequently arises is whether they are really different rationalities: don't scientists have always had the same beliefs about good reasoning? So the monism/pluralism distinction can be made at the descriptive level of the normative level. The proper description of epistemological ideas will either show that people have only one unitary view of what good reasoning is, or it will show that their epistemological beliefs importantly vary across individuals, time and cultures. We could call these (ideal-type) positions *descriptive normative cognitive monism* and *pluralism* respectively, since they are about the proper description of norms about cognition. The Strong Programme's methodological relativism (see also sect. 3.1.2) can then be formulated as being the methodological assumption that descriptive normative cognitive pluralism may be true: there may be different norms of good thinking that are relative to communities, and these norms and their history are worth studying. In particular, scientific developments of the past may have called on different principles of good reasoning to justify their truth claims. Methodological relativism opposes the presumption that there can be only one way to reason properly (normative cognitive monism), and that this way must have been used in the history of science, since science is essentially characterised as being a process of good thinking. When retracing what happened in the history of science, it is not sufficient to apply one's own and presumably unique norms of good reasoning to the historical situation under investigation.

Sociologists of scientific knowledge have argued, with case studies, that the history of science has known different epistemological norms, that different scientific domains can subscribe to different norms of good reasoning, and that the norms evolve as science evolve. It is argued in particular that competing theories may encompass their modes of justification, their principles of good reasoning, so that there is no common ground for comparing their epistemic virtues. This provides one line of argument for the incommensurability thesis (the other being grounded in semantic holism), and goes against the belief that scientific controversies are eventually resolved when sufficient data is gathered enabling picking the right alternative theory by applying unique norms of good reasoning (maybe the most subtle example of this view is Galison 1987). Did scientists have and did they use different epistemologies, or did they use only one? Again, the question is a matter of degree: how similar are the epistemologies used and developed in the history of science? The reason why I advocate methodological relativism – with the Strong Programme – is that the question of similarity is better asked when one does not already assume that they are similar. One of the arguments I will develop in this paper is that there are reasons to believe that these epistemologies are different from one another, but similar to a great extent because they are all produced by humans endowed with human minds. In order to pursue that argument, I need first to argue in favour of descriptive cognitive monism. This is what I will do in the second section of this chapter. Below, I show that methodological relativism is compatible with descriptive cognitive monism.

4.1.2 From minds to epistemic norms

What is the relation of methodological relativism with the views on lower levels listed by Stich? Since methodological relativist acknowledges that science has used and developed different epistemologies, it is natural for her to espouse 'normative cognitive pluralism', or simple relativism. This is not, however, necessary: the methodological relativist can acknowledge that different epistemologies have been used in the history of science without having the further belief that these epistemologies are true epistemologies; he can believe that only his own epistemology is the right one and that all others are flawed, and thus subscribe to normative cognitive monism. For my purpose, there is no obligation to take position on this matter. Let me just note that the recognition that different ideas of good reasoning have enabled scientific development leads to have a certain modesty with regard to one's own ideas of what is good thinking, and that methodological relativism with normative cognitive monism may be more difficult to defend because it implicates that most scientific thinking that happened during the history of science was not good thinking.

What is the relation between methodological relativism and descriptive cognitive monism and pluralism? The intuitive idea seems to be that methodological relativism fits best with descriptive cognitive pluralism: if there are different norms of good thinking, then there must be different ways to think in the first place. The existence of different norms of reasoning is easily explained when one subscribes to descriptive cognitive pluralism. For the relativist of the old tradition in cognitive anthropology, the explanation is given by the impact that social norms have on mental cognitive processes. People's ways of thinking is framed by their culture with the consequence that they conform to its norms. The problem with this explanation is that it is not psychologically plausible: the mind is not a tabula rasa upon which cultures write. The impact of culture on ways of thinking is an empirical question that obviously requires a detailed and complex answer giving due roles to what is learned and what is innate. Of course, the thesis that different people receiving similar stimuli at a time t will conclude the same thing and form the same beliefs is untenable. It is falsified by the simple fact that people disagree on the interpretation or way to act when faced with similar stimuli. It does not even hold for organism with much simpler cognitive apparatus than ours: for instance, rats put in front of some food may react differently depending on their past experience with food with similar smell. At a minimum, one must recognise that past experience can influence present cognitive processes through activation of memorised representations. More surprisingly, psychologists have shown that prior beliefs have a great impact on the formation of new beliefs even when these prior beliefs are discredited (Nisbett and Ross, 1980, chap. 8). This is of great relevance for the understanding of scientific beliefs formation, where theories are constantly challenged and eventually replaced by new ones.

The descriptive cognitive monism I want to defend is a monism about cognitive capacities. Of course, if the capacities that are said to be shared by all humans are only very broadly defined as a capacity to think, a capacity to reason, a capacity to be social and such general characteristics, then descriptive cognitive monism is just a truism, and it allows the possibility that humans in fact think very differently one from another. But if the capacities that are said to be shared by all humans are sufficiently specified, then descriptive monism gets some empirical content. The cognitive processes available to all humans are said to be very similar. For instance, perception of colours is said to be constrained by the human cognitive apparatus rather than by previous cultural input on colour names (Berlin and Kay, 1969): the available cognitive processes for colour perception are shared by all (normal) humans and do not vary with colour terms. Descriptive cognitive monism can be asserted also for capacities that are closer to reasoning than to perception. For instance, it is assumed that all humans attribute beliefs, desires and intentions to others in order to account for their behaviour. This ability is called theory of mind and is shown to involve complex cognitive processes implicating representations of others' representations (for a discussion on cultural variations and universals of the theory of mind, see Lillard 1998).

The rest of this chapter will show why and how descriptive cognitive monism of capacities is compatible with methodological relativism. One simple idea is that when different people with different epistemological beliefs form new beliefs on the basis of some new data, they may use different cognitive processes because they choose these processes with the intention that they comply with their ideas of what good thinking is. This rely on the basic observation that prior beliefs can influence the formation of new beliefs: epistemological beliefs do influence the formation of new beliefs in cases where people think reflexively, i.e. when, as often in scientific cognition, they pay attention to the quality of their thought processes. But how and why do people grasp and obey the epistemological norms of their communities? Here are two suggestions:

- Theory of mind, the cognitive ability to attribute beliefs, intentions and desires to others, is an important means individuals employ for anchoring themselves in a specific culture. Because individuals know how the others want them to think, they will conform to it in order to achieve their own goals. So, rather than understanding epistemic norms as explicit cognitive obligations set by some Platonic realm, people most often understand and answer others' expectations (e.g. the teacher's expectations). Epistemic norms then appear as macro-social consequences of people conforming and reproducing the norms by answering epistemic expectations.

- If people can conform to epistemological norms, then these norms must be somewhat adapted to the cognitive abilities of the people. This adaptation is not surprising, since these norms are the output

> of people thinking with similar cognitive abilities. The human cognitive apparatus importantly constrains the production of cultural epistemological norms.

At this point, what must be explained is why people with similar cognitive capacities would develop different norms of good thinking. Why does similarity of cognitive capacities not imply similarity of epistemologies? The answer I want to develop is that it does to a certain extent, but that historical and cultural contingencies are important factors in the constitution of beliefs about good thinking, and that differences in epistemological beliefs across communities are accountable with this historical factors. Fortunately, research in that direction is well advanced with the work of sociologists of scientific knowledge, who have pointed out socio-historical processes at the origin of judgements of rationality.

It is because theorists of rationality have ignored the complex relations between the nature of the human mind and cultural production that they have tended to associate either descriptive cognitive pluralism with methodological and epistemic relativism or descriptive cognitive monism with rationalism. For most rationalists, the culturally implemented norms of good reasoning are nothing more than the cultural image of the mind's rationality. For most relativists, the mind's processes are nothing but the psychological implementation of the culturally elaborated norms of reasoning.

The goal, however, is to understand the impact of the cognitive processes, as described by cognitive scientists, on the cultural norms that constitute 'rationalities'. The sociologists' data are approvals and disapprovals of reasoning; it is social normative behaviour which constitutes cultural norms. How does this social behaviour relate to human cognitive ability? Reducing social norms to cognitive abilities cannot be done without a thorough description of the causal chain that allows cognitive abilities to be 'reflected' or 'implemented' in social norms. Conversely, reducing cognitive abilities to cultural norms is not warranted without serious psychological studies. The intellectualist position is a step away from these simplistic reductions: faced with ethnographic reports about beliefs and reasoning not conforming our own norms of good reasoning, 'intellectualist' social scientists and philosophers (Horton, 1970; Lukes, 1982) have argued that these apparently irrational beliefs can be seen as rational when one provides a sufficient analysis of the context in which these beliefs have been formed. The differences of beliefs are explained with the differences of available evidence on the basis of which people think. Rational reconstruction can be seen as the application of the intellectualist position to the history of science. In this view, scientific beliefs differ from indigenous beliefs only because the former is based on much more evidence, much in the same way that past scientific theories are explained as differing from present theories because scientists of the past did not have the data we now have. The main difficulty, for the intellectualist approach, is that they are numerous beliefs – especially religious beliefs - that are supported by no evidence at all. Sperber's ethnographic example (1982), in his brief criticism of the intellectualist position, is Fataleka's belief that the earth is the fifth of nine parallel layers where crocodiles are in the seventh layer and flutes in the fourth. Moreover, the intellectualist approach does not explain why western science have managed to gather more and better evidence than other cultures, letting unanswered the question of the differences of beliefs. One could wonder if the intellectualist explanation applies to differences in epistemological beliefs: culturally variable categorisations of what is and what is not rational would then depend only on two phenomena: first, the implementation of our cognitive abilities, which are universal and rational, and, second, on the input to which these abilities are confronted. Such an account of epistemological belief variation oversimplifies the causal chains that produce norms of rationality. What importantly enters the account is the role of transmitted knowledge and transmitted norms of good reasoning. The intellectualist position is certainly on the right track, but gives too little importance to transmitted knowledge and to social phenomena in general.

Scientific knowledge production is the archetypical rational activity. But what is a rational activity? If rationality is simply described as the essence of human nature, then the fact that people do act in accordance with norms of rationality is certainly not surprising. What needs to be explained, on the contrary, is the existence of error. Against this long philosophical tradition, the framework henceforth described requires empirical investigation of the facts that enable people to be rational.

Science is a highly normative practice that obviously strongly draws upon cognitive abilities. On the one hand doing science requires obeying the rules set by the community of scientists. What counts as science is the fruit of the collective decision of scientists. On the other hand, doing science thoroughly involves our intellectual competencies, our feeling of certainty, our intuitions, and thus our individual capacities. Scientific production is thus conjointly constrained by social norms and cognitive competencies. The principles of scientific developments lie in the interaction of these two factors.

4.2 THE RATIONAL FOUNDATIONS OF RELATIVISM

This section develops two interrelated arguments: the first argument is an illustration of the point shown above: that debates over rationalism have mixed the empirical points about the nature of the human mind and the epistemological point about standards of good thinking. The presumed opposition between Sperber as a rationalist and the sociologists of the school of Edinburgh as relativists is therefore de-dramatised. This first argument, which pertains to the history of ideas in the social science, is a preliminary step for my main goal: putting both theoretical frameworks – the sociology of knowledge of the Strong Programme and the cognitive anthropology of Dan Sperber – at work for a naturalistic understanding of scientific thinking. I thus show that descriptive cognitive monists have developed good accounts of how beliefs may come to differ across cultures in such a way that they appear to lack epistemological grounds when seen from another culture. Whence the provocative title of the section: rationalism-contra-empiricism can explain why relativism-contra-rational-reconstruction is psychologically gounded.

4.2.1 Sperber's rationalism and methodological relativism

In 1982, Hollis and Lukes edited a book opposing relativists and rationalists. Among the authors were Bloor and Barnes, in the camp of radical relativists, and Sperber, in the camp of strong rationalists. It may therefore appear foolish or impossible to use Sperber's cognitive anthropology with Bloor and Barnes sociology of science. It is, however, both possible and desirable to do so; and this without compromising or diluting any of the claims of the apparently opposed camps. This is possible because the authors have different target for their criticism than each other; their argument illustrate the ambivalence of the rationalist and relativist positions.

Barnes and Bloor (1982) argue in favour of methodological relativism, which they back up with the view that they are indeed different norms of reasoning that have been held during the history of science and across cultures. As methodologists, Barnes and Bloor argue in favour of methodological relativism; as social scientists, they argue that there exist different ideas about what good reasoning is (descriptive normative cognitive pluralism); as philosophers, they argue in favour of normative cognitive pluralism. In addition, Barnes and Bloor are very much open to the psychological claim of descriptive cognitive monism. They say:

> What else is there to do then but to turn to causes for an answer to the question of the widespread acceptance of deductive inference forms and the avoidance of inconsistency? A plausible strategy is to adopt a form of nativism: the disposition arises from our biological constitution and the way the brain is organized. Such a move, needless to say, gives no comfort to rationalism: epistemologically, to invoke neuronal structure is no better than to invoke social structure; both moves seek explanations rather than justification. And for this very reason nativism is perfectly compatible with relativism.

Barnes, Bloor, and Henry (1996) consecrate the first chapter of their book to an analysis of the psychology of perception. The analysis is not a sociological analysis of scientific knowledge production; they authors consider the available theories (esp. the debate between Churchland and Fodor) as sociologists of knowledge interested in the cognitive foundations of knowledge production. Their conclusion is that the theory that asserts that perception is *not* or little influenced by beliefs is the most probable theory (i.e. they favour Fodor's contention). They also add a further reason for taking this psychological theory as true: methodological prudence prescribe not to help sociological claims with the most favourable psychological assumption that perception is theory-laden (p. 13).

Sperber (1982) argues against the view that people of different culture live in different worlds. This view, held by many anthropologists in the 70's, explains apparently irrational beliefs by the hypothesis that people from other cultures have a different experience (phenomenology) of the world. Because they 'feel' the world differently, people from other cultures generate radically different beliefs. Against this view, Sperber positions himself in favour of a nativist universalist position, which asserts that human cognitive abilities are to a large extent biologically determined and universally distributed among humans of every culture. Consequently, the phenomenology of any human is strongly constrained by the human biology and bound to be similar from one individual to the other. Sperber contrasts the strong similarity among the lived world of humans with the dissimilarity that may occur between different species endowed with different cognitive abilities. Sperber's arguments come from psychology: the relativists' assumptions about the working of the mind are implausible. Yet, as an anthropologist, Sperber does recognise that beliefs vary importantly across cultures. The differences are important enough to make others' beliefs appear irrational, i.e. as if produced by radically different cognitive processes than ours. In fact, his work in theoretical anthropology consists in showing how the human mind, as constituted of universally shared cognitive abilities, can constrain the content of culture *and*, at the same time, produce cultural diversity (Sperber and Hirschfeld, 2004).

Thus he lets open the possibility that there are different epistemological beliefs across cultures. It is hard to think that any contemporary anthropologist could have the conviction that epistemological beliefs *must* be the same across cultures. Remember, indeed, that epistemological beliefs are implicated in beliefs about the world. For instance, believing in sorcery is, among other things, believing that explanations implicating sorcery may be warranted; believing in sorcery is believing that thinking in terms of sorcery may be good thinking (although the latter belief is not necessarily explicit: only sorcerers or epistemologists will develop explicit metacognitive principles). The negation of methodological relativism is held only in the context of science studies, where philosophers still hope that science can be minimally characterised as essentially rational – with fixed standards of rationality and good scientific thinking.

4.2.2 Same world, same minds, but different beliefs . . .

In this rationality debate, Bloor and Barnes are in the context where rational reconstruction is still the orthodox method in the history of science and contest the ground of sociological explanations of scientific beliefs. Sperber does not address philosophers of science, but anthropologists. He is in the context where anthropologists use tailor made ad hoc psychological theories to explain away the existence of apparently irrational beliefs. This difference of context explains much of the difference of rhetoric – claiming to be relativist versus a rationalist and reciprocally – in spite of the similarity of their empirical claims. This similarity, however, appears when they attempt to solve similar problems. For Sperber, the challenge is to explain the great variety of cultural beliefs in spite of the similarity of cognitive means for understanding the world. For Barnes and Bloor, the challenge is to explain scientific controversies, characterised as situations where at least two sets of scientists form different beliefs, in spite of the fact that they have access to the same data and are supposed to have similar cognitive capacities. In order to explain differences of beliefs, both Sperber and Barnes and Bloor avoid appealing to differences of cognitive capacities and show that the relevant variable is the cultural and historical context. They are descriptive cognitive monists and at the same time unsatisfied by the twin approaches that are the intellectualist position and rational reconstruction. Last but not least, they both want to develop strongly naturalistic, causal explanations of why people hold the beliefs they do.

When Sperber attempts to explain apparently irrational beliefs, he factors in, not only the difference of input from the phenomena the beliefs are about (as an intellectualist), but also the social context and the deferential behaviours that are caused by the differences of social status. Barnes and Bloor insist that both parties of a scientific controversy have, except evidence to the contrary, similar cognitive abilities and do perceive the phenomena investigated in the same ways. From this, Barnes and Bloor concludes to the necessity of appealing to scientists' different cultural backgrounds, individual histories and interests, in order to account for the differences of scientific choice. But how can the social context make a difference on people's beliefs if they are all in a similar world and endowed with similar cognitive abilities? Sperber and the sociologists of Edinburgh answer with theories about cognition: they both make a distinction of types of cognitive processes.

Sperber distinguishes between factual and representational beliefs, or, in a later text (1997a), between intuitive and reflective beliefs. *Intuitive beliefs* are representations stored by the organism and freely used as premises in practical and epistemic inferences. We do not reflect on the way we arrived at intuitive beliefs or their specific justification, but simply hold and use them; they are grounded in perception and spontaneous and unconscious inference from perception. Intuitive beliefs constitute our phenomenology, i.e. the way we experience the world. Archetypically, intuitive beliefs are acquired by the direct impression of the world on our senses and its consequent incounscious processing by the brain. I see a tree in front of me, so I have the intuitive belief that there is a tree in front of me. *Reflective beliefs* are characterised as being beliefs that are mediated by a representation embedded in the representation that makes the reflexive belief: they are of the form V(R) where V is a validating epistemic evaluation (paradigmatically, 'it is true that') of R. Thus, reflective beliefs are intuitive beliefs about the truth of some representation, where the meta-representational comment (the content of V) provides a validating context of the embedded representation (the epistemic status of R). Typically, reflexive beliefs are acquired by thinking about linguistic representations: someone I trust tell me that there is a witch in the mountain, so I believe that "there is a witch in the mountain" is a true representation. I have a belief about the world, but this belief is reflexive because it is mediated by my thinking about a linguistic representation as true. In some cases, adds Sperber, reflexive beliefs are mediated by a representation whose content (i.e. what it says about the world) is not clearly understood; the representation is *semi-propositional*. For instance, I can believe that there is a witch in the wood without fully understanding what a witch is (e.g. I may not understand how and why spells work). Sperber purports that all apparently irrational beliefs are reflective beliefs whose embedded representation has no clear content. First, "beliefs reported by anthropologists are representational" because "they are cultural beliefs, i.e. representations acquired through social communication and accepted on the ground of social affiliation" (1982, p. 175); second, if they appear irrational to others, it is because their convincing appeal implies, to a large extent, a deferential attitude towards those who communicated the embedded representation (e.g. my father, the chief of the village, *whom I trust*, told me: 'there is a witch in the wood'). Why do humans hold such reflexive beliefs? Because they have the ability to process representations as representations: they can think about representations as having meaning. With this ability, a meta-representational ability, they can gather information from others, through communication. They can progressively learn from others and specify the meaning of their own reflective beliefs.

Barnes, Bloor, and Henry (1996) start with the psychological assumption that perception is independent from prior beliefs and show how prior beliefs nonetheless influence scientific belief formation. The explanation is especially called for in cases where the diversity of scientific beliefs occurs in scientific controversies, that is, cases where intellectualists, or rational reconstructivists, run short of explanatory resources. Scientific belief formation is influenced by prior beliefs because, Barnes et al. purport, scientific beliefs are not direct perceptions of how the world is, but theoretical interpretations of what our perceptions reveal about the world. Interpretation, they argue, intervenes pretty soon in scientific activity: observation reports implicate presumptions and assumptions that are often of a highly theoretical character. Barnes et al. illustrate their claim with the study of Millikan's experiment for measuring the fundamental unit of electric charge.

They show how Millikan's conclusions were sensitive to how he treated his data – a treatment that implicated judgement and selectivity. Millikan's interpretive procedure called on the resources of a scientific tradition, which is, as all interpretive traditions of science, "largely inherited from others, shared with others, validated by others and sustained in the course of interacting with others" (p. 26). The pervasive role of interpretative scientific tradition is shown by the simple fact that a lay man could not even make sense of Millikan's experimental apparatus. But the decisive role of the traditions and their locality is illustrated with the Millikan-Ehrenhaft controversy. In this controversy, Millikan draws on a realist stance about electrons, already hinted at, but not taken seriously, by Maxwell when talking about molecules of electricity. Ehrenhaft maintains on the contrary that there is no elementary electrical charge and relies on the assumption that what matters, in physics, are law-like regularities detected empirically rather than the postulation of invisible theoretical entities. He was an empiricist along the lines of Ernst March. Interestingly, the interpretive traditions thus implicated clearly epistemological stances. The example eventually shows that "any attempt to interpret the world theoretically must, at some stage, impose its categories and meanings, and at this point, inevitably, it will be both risky and dogmatic" (p. 24). It is risky because it is bound to rely on an interpretive tradition, which must be, to a significant extent, dogmatically accepted in order to make sense of the data. Barnes et al. conclude: "the mind of the individual scientist is the point of contact between our physical environment and our social environment. Interpretation is where nature and culture come together" (p. 28).

Sperber and Strong Programmers' hypotheses about how culture gets in, in belief formation, are quite similar. More precisely, Sperber can be taken as fleshing out the psychological assumptions of the interpretation phenomenon called upon by sociologists of the Strong Programme. He says "it is arguable that much of culture, from religion to science, is made of reflective concepts and beliefs". Thus, scientific beliefs are reflective beliefs; they take the form of V(R), where R is a scientific theory or hypothesis, and V its validating context. Here is what Sperber (1997a, p. 77) says about scientific beliefs:

> Many well understood beliefs are reflective beliefs, paradigmatic examples being scientific beliefs. Some of the concepts that are used in scientific claims are well-understood by scientists but the may remain beyond the reach of their intuitions. These are concepts that scientists can indeed think with, but, in most cases, only by thinking about them, that is, only reflectively. Typically, the validating contexts of beliefs containing such scientific concepts are not (for competent scientists) in the form of a reference to an authority, but in the form of an argument or a demonstration. Such arguments and demonstrations are not of a kind delivered by spontaneous inference, and must be reflected upon in order to see their force.

Likewise, when talking about interpretation, Barnes et al. intend to contrast it with beliefs acquired through direct perception, i.e. intuitive beliefs and insist that interpretive procedures of belief formation imply interpretive traditions. These interpretive traditions, of course, provide the validating context. Whether the validating contexts of competent scientists involve or not reference to authority is a matter that is currently investigated, and debates in social epistemology question the conditions under which reference to epistemic authority is justified (see the work on trust and testimony, e.g. Hardwig 1985, 1991). However, arguments and demonstrations are traditions and they need to be selected by scientists. Often, this selection goes without saying for scientists; sometimes, as in the cases of controversies, the fact that scientists choose their interpretive tradition is made manifest by the difference of choices. Millikan and Ehrenhaft did not select the same traditions. Another property of scientific beliefs appears in both Sperber's and the Strong Programme's works: it is the fact that scientific concepts can always be redefined in order to fit some explanatory trend or purpose. With the Strong Programme, this possibility is strongly theorised under the label of 'meaning finitism'. But meaning finitism is essentially based on remarks concerning the logical and epistemic status of terms, as having been used in a finite set of situations; it says little on the psychological status of scientific concepts. Sperber's theory of intuitive and reflective beliefs points at an answer (1982, p. 170):

> if one finds oneself holding two mutually inconsistent ideas and reluctant to give either, there is a natural fallback position which consists in giving one of them a semi-propositional form. This occurs, for instance, in scientific thinking when counter-evidence causes one, instead of rejecting the theory at stake, to search for a new interpretation of it by making some of its terms open to redefinition. As long as this search is going on, the theory is in a semi-propositional state.

Sperber argues that some concepts (as mental entities), scientific concepts in particular, can change status from intuitive to reflective and from reflective to intuitive. For instance, we have in the course of development a concept of weight that is used in spontaneous, rather than deliberate and conscious, inference. This intuitive concept is used when we catch falling objects, for instance. However, this concept is questioned when it comes into conflict with the concept of mass. The concept of weight becomes reflective because inferences that implicate it call on reflective beliefs such as "'weight' is different from 'mass'". The concept of weight is somewhat put into quotation mark so that its meaning be determined by the reflective beliefs where it is implicated, rather than by direct intuition. Perhaps sufficient reflective mastery of the changed concept makes it become intuitive again. It is arguable that scientific beliefs contain a significant number of semi-propositional beliefs, since scientists always investigate the actual consequences of their own theories. And there is an empirical question as to when and how concepts change their status between reflective and intuitive . The point, however, is that there are psychological grounds for understanding meaning formation in scientific development, and that these psychological grounds need not postulate an unwarranted plasticity of the mind. On the contrary, Sperber would lean towards the language of thought hypothesis: there would exist a set of innate concepts out of which thinking is made.

4.2.3 A naturalistic look at reasons

So, eventually, do Sperber and Barnes and Bloor agree in the rationality debate? I have shown that they actually have many claims in common. I would even say that their empirical claims are fully compatible. However, their positions with regard to normative cognitive monism or pluralism seem to differ. Barnes and Bloor argue that there is no principle of good reasoning that is above and separated from the particular implemented ideas of good reasoning. So, there is no epistemology above scientific practices and reasoning, but scientific practices and reasoning include local epistemologies. Their methodological relativism is backed up with a pragmatic relativist epistemology. By contrast, Sperber (1982) calls on intuitions about what good reasoning is in order to justify his rationalist position: after explaining that deferential behaviour is at the basis of most apparently irrational cultural beliefs, he evaluates the rational status of this behaviour: “when all the members of your cultural group seem to hold a certain representational belief of semi-propositional content, this constitutes sufficiently rational ground for you to hold it too” (p. 177). In this sentence, ‘rational’ does not refer to the actual structure of the brain but to universal standards of good reasoning. On the basis of his anthropological experience, Sperber asserts that people defer to other people *de* *facto*, and he further *judges* that this is rational. However, beyond these epistemological stances, lies, I think, the same motivation: showing that people from other cultures who hold apparently irrational beliefs and scientists who developed wrong theories are not dumb, and their thoughts and behaviour is not out of reach of scientific explanation [1] . What I have shown, however, is that this difference in epistemological position does not really matter for the naturalistic investigation of scientific knowledge. Indeed, both Sperber and the sociologists of Edinburgh are strong advocates of naturalism in the social sciences. For Barnes and Bloor, naturalism is the reason why they reject dualist explanations for belief formation, according to which false beliefs are explained with causes, and true or scientific beliefs are explained with reasons – understood as only epistemic entities. The symmetry principle of the strong programme consequently prescribes studying the causes of beliefs independently of their semantic or epistemic properties. With scientific practices, where reasons are explicitly invoked as causes of beliefs, the sociologists of Edinburgh stand firm: epistemic properties have no causal power by themselves, since they have no natural status; reasons as causes cannot be but beliefs, but these beliefs must be sufficiently partaken so as to be accepted as reasons by others; eventually, it is credibility that gives causal power to claims of validity, and it is argumentative appeal that change mere beliefs into accepted reasons. Likewise, Sperber, as a naturalist social scientist, is interested in reasons only because, and insofar as, they have causal power on the decisions of social actors. In a criticism of rational choice theory in sociology (1997b), he says: "le seul type d'individu manifestement naturel, ce n'est

[1] A sociological analysis of their scientific beliefs would certainly point out the different scientific cultural backgrounds against which they reacted. Sperber was educated in anthropology, where relativism was developed as a permission not to solve the problems met by symbolic and intellectualist approaches – "people just think differently". Bloor, as a representative of the school of Edinburgh, was educated in the philosophy of science, where foundational epistemology was collapsing. In 1966, Sperber was in Oxford studying social anthropology and Bloor was in Cambridge studying psychology. It seems just natural that the former pursued his investigation on the psychological foundations of cultural beliefs, and the latter pursued his research on the cultural foundations of scientific belief formation.

ni le sujet, ni l'agent, et encore moins l'agent rationnel, c'est l'organisme [2] ". Thus, "les raisons nous intéressent quand elle jouent un rôle causal, et elles ne nous intéressent a priori ni plus ni moins que tout ce qui joue un rôle causal [3] ". He then pursues with a naturalistic definition of 'reason':

> Parmi les états mentaux, on a les croyances et les désirs qui suscitent les actions, autrement dit, les raisons des acteurs. Appeler ces états mentaux "raisons", c'est mettre en valeur le fait qu'ils relèvent aussi d'une évaluation normative, d'un jugement de rationalité (tout comme nommer certains états mentaux "connaissances" implique qu'il relèvent d'une évaluation normative, cette fois épistémique). Les raisons n'en sont pas moins des causes parmi d'autres [4].

The normative evaluations, would continue Barnes and Bloor, consist in calling upon the social norms of one's community, i.e. the accepted criteria of truth and rationality of one's culture. This sends us back to Sperber's cognitive account of epistemic evaluation as meta-representations generating reflective beliefs.

Scientific thinking is reasoning with scientific ideas. Scientific thinking makes extensive use of the validating context, or the interpretive tradition,

[2] Naturalism should also be at odds with individualism in the strong sense, for only organisms, and not persons, or rational actors, are manifestly natural entities. (My translation, on the basis of Sperber's online translation, which is available at http://www.dan.sperber.com/individ.htm - I have completed it because Sperber's translation shortens the French source and do not include assertions that are relevant for my purpose.)

[3] from a naturalistic point of view [...], reasons should be of interest to us not qua reasons, but qua causes among other causes. They are not a priori more interesting than other causes. (Sperber's translation).

[4] Among mental causes, one has beliefs, desires, and practical syllogisms leading to actions. Beliefs, desires, and practical inferences may be described as "reasons". To do so is to draw attention to the fact that these mental states and processes can be evaluated from a normative point of view: they are open to a judgment of rationality (likewise, naming 'knowledge' some mental states implies an epistemic normative evaluation). Nonetheless, reasons are causes among others. (My translation, on the basis of Sperber's translation).

in order to think about the phenomena investigated. Doing so means processing representations that are reasons, because they are part of a framework agreed upon by a scientific community. Scientific thinking is also reasoning to the extent that new representations produced are aimed at constituting reasons for further scientific thinking. New representations are produced, which acquire their scientific status when they are recognised as such by the scientific community. Thinking scientifically, to say it bluntly, is using and enriching scientific knowledge. Sperber and sociologists of scientific knowledge specify the meaning of this truism; it consists in specifying the rational foundations of relativism, i.e. the cognitive mechanisms out of which different norms of reasoning emerge.

Descriptive cognitive monism is not refuted by the diversity of cultures and by the historicity of science. There is, on the contrary, a good account of this diversity describing the cognitive abilities that factor in cultural knowledge in belief formation. There are two reasons why scientists of science should be descriptive cognitive monists: the first one is because it provides a good account of the nature of the mind; the second reason is that this good account has real implications on the content and history of scientific knowledge. I develop on these reasons in the next chapter.

Chapter 5

Nativism for science studies

In the previous chapter, I have maintained that the idea of 'natural rationality' is compatible with variability of norms for good reasoning. I have begun to provide a picture of the human mind as endowed with interpretative ability—specified as the ability to meta-represent intuitive representations. In this chapter, I enrich this picture and assert that the mind is made of many other evolved cognitive abilities. I thus expose some arguments in favour of nativism — the assertion that human cognition is constrained in significant ways by the human genetic endowment—and use results from cognitive science, which have described some of the capacities with which humans are endowed.

The general goal is still to see the consequences of this view of the mind for social studies of science. In particular, I check the compatibility of the nativism held by evolutionary psychologists with the social constructivism of the Strong Programme. The compatibility claims are not just theoretical exercises in the philosophy of science studies: they are intended as prolegomena for the development of a more comprehensive picture of scientific activities and the evolution of scientific knowledge. If some cognitive abilities put at work in scientific cognition are not the mere reflected image of cultural facts, then psychological considerations can be of great relevance to explaining science and its content. Nativism is important for a science of science because the innate properties of the mind have some causal effect on the content of science.

The chapter is organised in four sections. In the first section, I show the complexity of the notion of innateness. I thus can defend a nativism that goes beyond the innate/acquired distinction and that take into account recent studies of the development of organism. The second section shows the consequence of nativism for naturalistic account of scientific thinking and scientific evolution. The third and fourth sections question the compatibility of nativism with the social constructivism of social studies of science and theories about scientific evolution. The main stake of the chapter is to question the role of innate factors of cognition in scientific thinking.

5.1 NATIVISM AND COGNITIVE ABILITIES

5.1.1 SOME COGNITIVE ABILITIES WITH INNATE BASES

Renewed interest in nativism, after the behavioural school of thought in psychology, came from Chomsky's argument showing that language could not be acquired on the basis of the utterances heard only. The grammar of a language, in particular, is always underdetermined by the finite number of utterances that children can hear. Chomsky concluded that human must be endowed with an ability to learn natural language: this ability constrain and enable learning, it embodies innate knowledge of grammatical principles that apply to all natural languages. This argument is known as the poverty of the stimulus argument, and is used in other domains of knowledge. Evidence of the innate basis of several cognitive faculties has been especially gathered in developmental psychology, which found out that young children could perform tasks at ages where the faculty required for the performance could not possibly have been acquired (see Spelke, 1998, about the role of developmental psychology in informing the debate between nativism and empiricism).

Numerous cognitive scientists have argued that the mind is made out of domain specific cognitive abilities which work on specific inputs, such as visual objects, the behavior of people or linguistic input(see, e.g. Barkow et al., 1992; Hirschfeld and Gelman, 1994; Sperber et al., 1996; Carruthers and Chamberlain, 2000; Carruthers et al., 2005). For instance, very young children assume that objects are connected bodies that maintain their connectedness over motion (naïve physics). Similarly, children and adults assume that people have desires and beliefs that cause their behaviour (Theory of Mind). In addition to studies on not yet acculturated infants, studies in cross-cultural psychology have shown the existence of universal features in folk knowledge that suggest the existence of underlying universal cognitive abilities (Folk biology; Atran 1998). The arguments in favour of the existence of such domain specific abilities are based on results in developmental psychology (e.g. the work of E. Spelke on naïve mechanics), evolutionary considerations and, to a lesser extent, cross-cultural psychology (e.g. the work of Atran on naïve biology). It is widely acknowledged that perception is carried out by a set of innately specified cognitive modules, and that knowledge of languages is based on a specific 'Language Acquisition device'. Here is a list of other well documented human modular abilities (taken from Sperber and Hirschfeld 2004 — which also contains bibliographic information for each capacity):

Theory of mind	Capacity to interpret behavior in terms of mental states like belief and desire
Folkbiology	Capacity to sort living things in terms of their morphology and reason about them in terms of biological principles like growth, inheritance, digestion, respiration, etc.
Number	Capacity to distinguish collections of objects according to the (small) number of elements in the collection
Face recognition	Capacity to distinguish conspecific faces from other similar stimuli and to identify individuals by the specificity of their faces
Naive mechanics	Capacity to form consistent predictions about the integrity and movements of inert objects
Folk sociology	Capacity to sort conspecifics into inductively rich categories, membership which is based on (supposedly) shared intrinsic natures

5.1.2 HOW TO CHARACTERISE INNATENESS

I have specified that the domain specific cognitive abilities above mentioned have an innate basis. But what does that mean to have an innate basis? Intuitively, the concept is the opposite of acquired. However, there is no empirical criterion to tell apart what is innate and what is acquired. The criterion 'present at birth' is inappropriate because many properties that we consider as definitively innate appear later during development, such as hairs on the pubis. This is true for many properties of the organism: genes may determine the height of a person to be around 1m70 (which also depends to some extent on his individual history), but at birth, the same person with the same genes is 50 cm long only. Likewise, many properties of the mind can be genetically determined but phenotypically manifest only at the adult age. Changes that happen during the lifetime may be directly determined by the genes.

Genetic determination seems to be a good characterisation of innateness. But in fact, the criterion leads to *aporia*. This is because phenotypes are always the joint products of genetic determination and environmental causal effects during development. The environment is clearly a condition for the biological development of organisms, which need a normal ambient temperature, food that is appropriate, and so on. So genetic determination always under-determines the phenotype and cannot characterize, on it own, what is innate. The causal effect of development is pervasive. Genetic under-determination has been vehemently emphasised, when concerning human behaviour. Genetic determination does not seem to account for the flexibility and context dependence of human behaviour; it does not seem to make justice to the importance of ontogenetic development.

At the outset, the assertion that genes partly determine behaviour is so obvious that it is useless: cows don't hunt hares, and foxes don't chew grass; these differences in behaviour are caused by differences in genetic endowments. Likewise, human behaviour is specifically human because it is so determined by the human genome. On the other hand, the strong claim that directly associates genes with behaviour (the genes for being homosexual or mathematician) is just wrong.

Genes firstly determine the biological constitution of living organism, and it is only through this constitution that genes can determine behaviour. Of great relevance is the biology of the brain, which determines psychological events, which in turn determine behaviour. Specifying this causal chain allows pointing out where other causal factors, such as context or individuals' history and learning, come in. But acknowledging the pervasive causal action of non-genetic factor is to say that the notion of innateness-as-genetic-determination can never classify cognition and behaviour as innate or not. Genetic determination, however, can still be used as a more or less notion depending on how proximal are the genetic causes, on how much constrain they put on cognition and behaviour, or on how little variation there is in cognition when the environment variates. For instance, the important role of genetic determination is conveyed by such terms as ’maturation’, which is defined in opposition to ’learned’. In these cases, the pervasive role of the input is coined ’triggering’ which emphasises how little it contributes to the content of the knowledge or ability that has matured. Three types of causes can be distinguished: the initial state, which includes initial knowledge and processes, the incoming input, and the genes. Depending on the cognitive phenomena, the genes have strong direct causal effect, or its causal effect is small or indirect via the determination of the initial state. The difficulties to pinpoint the biological realisation of what is innate at the neurological level has been strongly emphasised by Elman et al. (1996) in *Rethinking Inateness*, who argue that there is no specific and fixed patterns of cortical synapses (as determinant of mind/brain activity) that would represent innate knowledge before activity.

Considering the above pitfalls, Sober (1999) proposes to ’salvage’ the concept of innateness-as-genetic-determination, while acknowledging the pervasive causal role of the environment, with the following definition: ”a phenotypic trait is innate for a given genotype if and only if that phenotype will emerge in all of range of developmental environments”. The criterion for innateness is then developmental invariance. Samuels (2002) notes that this definition does not work either, because it classifies too many things as innate. For instance, the knowledge that water is wet is acquired in all environments where organisms can live, although this knowledge is certainly learned and never qualified as innate. Conversely, innate structures that require the fixation of variables, such as grammar, lead to much variation in the phenotype, such as knowledge of diverse natural languages. I would add that Sober does not provide any guide—pragmatic or not—that would help the scientist to circumscribe the range of developmental environments that figure in the definition of innateness.

5.1.3 Evolutionary causal cycle: genes—neural device(s)—cognitive processes—behaviour—gene selection

Evolutionary psychology is a subfield of psychology which assumes that the brain is a biological organ whose function is to issue adaptive behaviour. It aims at specifying the consequences of this assumption on the principles of human cognition. In order to do that, it needs to explain the causal chain from genes to behaviour. In particular, cognitive abilities cannot be reduced to the genetic endowment that makes them possible: in addition to the genetic conditions of possibility, cognitive abilities are implemented by a biological apparatus, the brain, and develop and change during the organism's individual history. Development always has a role in the constitution of cognitive abilities. The cognitive apparatus of the human adult is always, for every aspect, the product of both evolution and development. The growing field of 'evo-devo' has pointed out that this is true of all phenotypes. The graph 5.1 pictures the pervasive role of the environment in the determination of cognition and behaviour.

And yet, evolutionary psychology does need some notion of innateness at the behavioural level in order to account for the differential fitness that different behavioural dispositions may have. How can adaptive behaviour evolves? There is a gene-behaviour causal cycle: the selective pressures operate on behaviour, but it is the genes that are selected. Imagine two antelopes chewing some leaves; a lion comes; one antelope flees away, the other one eats undisturbed ... until the lions kills it. The differential behaviour may stem from either difference in the individual history of the antelopes (i.e. contextual factors) or from genetic differences. In the latter case only, the multiplication of the 'hunted antelopes' above situation will lead to the selection of the genes that favour fleeing behaviour. While in each actual case of hunted antelopes, actual behaviour of antelopes are determined through both genes and individual history, the *repeated differential behaviour* is to be accounted by an innate disposition to flee. The abstraction of the individual history determinants is justified by the fact that natural selection picked out the genetic determinants only. In this way,

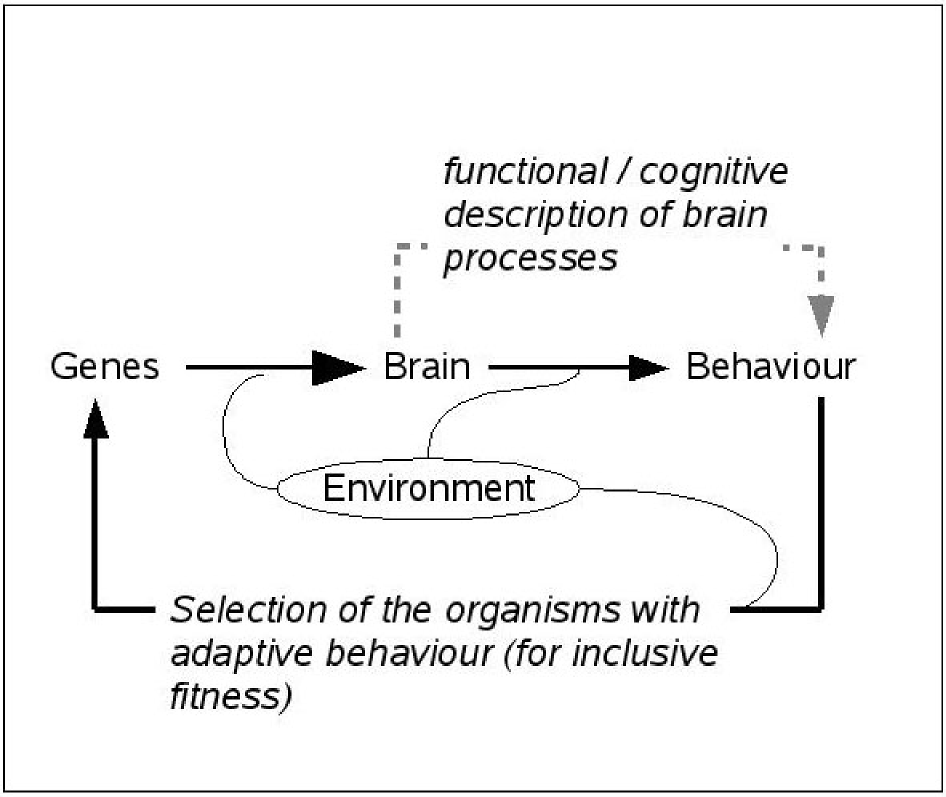

Figure 5.1: The causal cycle: genes—cognitive devices—cognitive processes—behaviour—selection, where 'cognitive devices' and 'cognitive processes' correspond respectively to functional description of the brain and of brain activity.

the effect of a set of genes that has contributed to the selection of the set can be qualified as innate. In other words, evolved properties are innate. This use, it must be noted, is restricted to species-typical traits, because it is at this level only that the selective operations of evolutionary history can be observed. It does not allow discriminating innate properties in one individual (e.g. autistic behaviour can be qualified as 'innate,' because it is largely determined by the genes of the autistic individual).

When can we say that a cognitive ability is an evolved, rather than an acquired ability? Evidence can come from evolutionary constrains, from developmental psychology, and experimental psychology—is the ability adaptive? Does it vary little across individuals, or across cultures? Does it appear before it could be plausibly acquired on the basis of input? Positive answers to these questions can function as evidence that the cognitive ability is evolved, and thus innate.

Research in evolutionary psychology is centred on the psychological and cognitive level of this causal chain from genes to behaviour, and one of the key questions is about the genetic determination of the constitution of the mind.

Cognitive psychologists have argued, with great empirical evidence, that humans are endowed with innate cognitive abilities. I have endorsed this claim and I have specified a way to understand what is meant by 'innate' in 'innate cognitive abilities.' In the rest of this part, I will show that these innate cognitive abilities have causal effects on scientific knowledge production. In my case study, in particular, I will show how the evolved capacity to understand magnitudes has had some consequence on the history of mathematics.

5.2 Why the structure of the mind constrains the content of science

5.2.1 Against the Blank Slate view of the mind

In their investigation about 'the psychological foundations of culture', Tooby and Cosmides (1992) advocate an "integrated causal model" for the social sciences. This model is first and foremost a naturalistic programme that promotes causal explanations in the realm of the social. Such a naturalistic programme, Tooby and Cosmides argue, requires building bridges between the social and the natural sciences. Of course, there is no question of eliminating the social sciences with the idea that natural sciences can do the explanatory job. But the goal is to understand what sort of causal events underlie social and cultural events and what natural phenomena make social phenomena possible. Tooby and Cosmides point out the importance of two bridges: between the social sciences and cognitive psychology, and between cognitive psychology and biology. The essential import of this approach is that biology sets constraints on psychology, because mental processes are implemented in a biological organ, and psychology sets constraints on the social sciences, because social phenomena are realised by interacting thinking beings The authors contrast their proposal with current practice in the social sciences. They accuse the "Standard Social Science Model" of considering the human realm and social order as outside of the natural causal world and to assume erroneously that the human mind is a like a blank slate. Tooby and Cosmides forcefully attack the theory that views the human mind as a blank slate and its modern counterpart, the mind as a general-purpose mechanism. Findings in developmental psychology have shown that infants have some knowledge of the environment that is manifest so early as not to be possibly acquired in development. Against the blank slate metaphor of the mind, humans are endowed with innate knowledge. Among the reasons for the rejection of the existence of a general-purpose mechanism is evolutionary plausibility: there is no plausible evolutionary history for the advent of general-purpose cognitive abilities. Also, general-purpose mechanisms are unable to solve the current problems humans solve easily and rapidly (c.f. also the frame problem). More recently, Pinker (2002) has forcefully defended the scientific account of the innate organisation of the mind against the 'dogma' of the blank slate.

The theory asserting the existence of domain specific faculties presented in the previous section is a viable alternative account of the mind as a blank slate or as a general purpose information processing device. These evolved abilities allow dealing with the input in a quick, cost effective and reliable way. In order to achieve this, the cognitive devices have to assume things about the world — they have content; they embody innate knowledge.

One consequence of this view is that cognition does not consist in the mere extraction and organisation of information from the world; rather, it involves cognitive capabilities that shape our understanding of the world according to their internal organisational principle. Our naïve and direct apprehension of the world involves cognitive construction. Tooby and Cosmides urge us not to stop here our investigations on cognition. Their Integrated Causal Model reveals the possibility and importance of linking these findings to evolutionary biology. The cognitive devices of the human mind are indeed themselves organs that have been selected during evolutionary history. A complete causal account of cognition and behaviour must therefore include an account of the evolutionary construction of the evolved information-processing device [1] The relevant aspect of this argument, for our purpose, is that one cannot explain scientific cognition in terms of some special scientific competency if this competency does not have a plausible evolutionary history. Ironically, a similar argument is developed by Latour (1986) who conclude that cognitive psychology is not relevant to science studies. By contrast, the conclusion I want to draw is that since scientific rationality as postulated by philosophers is not implemented in the mind, then scientific cognition needs be re-thought with the best theories in cognitive psychology. Moreover, one can expect that if the mind is not tailor made for science, then science must develop so as to 'fit' the nature of the human mind. In other words, scientific knowledge must be cognizable. Therefore the nature of the human mind can be expected to constrain the content of science.

[1] It is worth noting that the programme of evolutionary psychology put forward by Cosmides and Tooby and others, and which I present as convincing, is NOT aiming at finding presumed biological causes for socio-cultural inequalities. On the contrary, the research programme is first and foremost aimed at understanding the universal psychological traits of the human species.

5.2.2 SCIENCE STUDIES WITHOUT THE DOGMA OF THE BLANK SLATE

Cosmides and Tooby's attack, indeed, bears on Science Studies. The criticised view of the functioning of the mind is at the basis of the argument of both those who deny the role of cognition and those who deny the role of social interactions in scientific knowledge production. In each case the mind is but an empty bag that is either filled by the 'social' or through empirical observations only.

The latter view has been sustaining some form of 'direct realism' — the view that scientific truths are directly (without social bias) given to the mind observing the world. Maybe the most elaborate view of that sort has been proposed by Gopnik and Meltzoff (1997), who argues that the development of science is accounted for by the mere accumulation of data made available to the minds of scientists: data are processed by the mind's general-purpose cognitive abilities, which compile scientific theories. Not surprisingly, such a view has been criticised for not taking into account the social nature of scientific practice by even the most people opposed to the Strong Programme (e.g. Carruthers, 2002, section 3). Direct realism overlooks the determinative role of active experimenting (scientists are not just 'listening' to what the world says, they also choose which questions to ask), communication (scientists need convince others), and the inter-subjective control of scientific production that is supposed to guarantee objectivity. However, the fact that such views are attacked not only for their misconception of science but also for being wrong about cognitive processes opens new prospects for the analysis of the relation between cognitive and social phenomena in scientific knowledge production. Hutchins (1995) has noted that cognitive scientists have made the mistake to attribute to the individual mind processes that are in fact done by larger cognitive systems, sometimes including several people. Science is a case in point: it is thought that it could be the product of isolated scientists, as if one brain living long enough would have developed the same kind of knowledge as centuries of scientific disputes and confrontations of ideas The hypothesis that scientific thinking relies on evolved cognitive abilities provides a promising alternative, as it does not take science, a collective achievement, as the paradigmatic illustration of what minds can do. In the perspective of evolutionary psychology, mental abilities have a biological basis that has evolved as issuing adaptive behaviour; but doing science is not a behaviour that played a role in the evolutionary history of the genetic basis of mental abilities; science is much too recent and local. Scientific cognition not being already squeezed within scientists' mind, one can wonder anew about what, in the mind, makes science possible: take what abilities have probably evolved, what abilities account for day to day behaviour, what abilities seem to exist before scientific training is done; on this basis attempt to account for scientists' behaviour out of which science emerge. This is, for instance, the explicit method used by Carruthers (2002), when he attempts to relate the cognitive skills of the hunter — which have presumably been selected for — with the skills put at work in doing science.

The opposite 'social reductionist' view, which denies any determining role to our cognitive apparatus in the formation of scientific beliefs (Latour and Woolgar 1986, p. 280; Latour 1987, p. 247) is also at odd with the thesis that the human mind is endowed with domain specific abilities that significantly constrain cognition, and thus belief formation. Numerous works in social studies of science appear to be embedded in the Standard Social Science Model and its erroneous assumptions about how the mind works. Insofar as one admits that the scientist's mind is involved in scientific knowledge production, the *a priori* claim that the nature of the mind has no role in framing knowledge cannot be justified but by the assumption that the human mind is totally plastic, so that it does nothing but reflect the social structure or the context. The social reductionist view, insofar as it is committed to the thesis that the mind is totally framed by enculturation, is erroneous. Against 'social reductionism', scientists use their biologically endowed and already content loaded cognitive apparatus when investigating the world. Rather than just taking on current scientific theories, they put their cognitive resources at work to understand, remember and apply them. Thus the cognitive structure and innate endowment of our mind is always empowering and constraining our thoughts — whether the thoughts are meant to be scientific or not. In particular, scientists are always forced to rely and assess the relevance and meaning of their intuitions. Intuitions remain at the basis of scientific reasoning. One reason why scientific cognition cannot do without intuitive beliefs is that reasoning and argumentation must always stop at a point (this is tantamount to the Wittgensteinnian remark about rule following: at some point the rule must be applied without having a meta-rule explaining how to apply it). This let scientists in front of claims that must be accepted at face value. Scientists accept claims without arguments proving its truth for different reasons. One reason can be that the claims are upheld by some other scientists whom they trust. Another pervasive reason is that the claims are obvious as they are. Why is that so? The intuitive appeal of the obvious claims is the result of some mental cognitive processes. How dependent from cultural knowledge are the processes that provide intuitive appeal? This is a question that cannot be answered a priori. Probably, the answer depends from case to case. One sure thing, however, is that these processes are not independent from pre-wired [2] mental mechanism; in particular, they are not independent from the pre-wired aspects of perception modules (i.e. what we perceive is never fully determined by what we believe). I will further argue that they are not independent from conceptual abilities (i.e. mental mechanisms not directly dealing with perception) such as the domain specific abilities mentioned above

5.2.3 SCIENCE STUDIES WITH THEORIES OF EVOLVED COGNITIVE ABILITIES

The relation between the cognitive and the social in scientific knowledge production can fruitfully be re-thought along the following questions: What is the role of evolved cognitive mechanism in scientific practice? How can a species, which evolved as a hunter-gatherer species, do science? The in-

[2]'pre-wired' means that the genes have a significant role in determining the brain structures that perform the cognitive processes (see Marcus, 2004)

tegrated Causal Model shows the relevance of these questions by putting forward the results of developmental and evolutionary psychology, which assert that the mind is so structured that it heavily constrains the content of our thoughts — including, of course, our scientific thoughts. Now, if scientific thoughts are produced by some highly specified mental abilities, then there is important reason to think that these abilities did let their imprint on our scientific knowledge. We have the opportunity to analyse the implications of the nature of the mind, as it is discovered by psychologists, on the content of science. We have the opportunity to understand scientific history and scientific practices more thoroughly.

From the Integrated Causal Model, the Strong Programme in the sociology of science already has the 'Causal'; I claimed in chapter 3 that it should take the 'Integrated' too. Integration is not only necessary for exhaustive causal accounts; it can also allow rethinking the classical tenets of social studies of science. The idea is not to deny the normative, institutional and cultural nature of scientific knowledge, but to show the strong causal power of the structure of the mind on scientific practices and theories. Beneath cultural stability and diversity lay the universal, yet contingent, nature of the human mind, together with its richly endowed organisation and domain specific processes (Sperber, 1996a; Sperber and Hirschfeld, 2004). While sociologists of science have concentrated mainly on periods of scientific changes and controversies, the stability of scientific theories is also worth explaining for those who insist on their cultural contingency. There is the fact that once a theory is institutionalised, new scientists are culturally led to use it. But what are the mental resources that allow a scientist to apply a given theory or to work within a paradigm? What makes a scientific work be an application of a theory, i.e. be sufficiently similar with past applications but yet in a new context? What does the scientist have in mind when she uses or thinks about a theory? /citetgiere88 expounds an influential theory about the nature of the mental representations involved in believing scientific theories: he asserts that these representations are 'models' rather than linguistic axiomatic systems. My contention is that the nature of the human mind influences not only the format of the mental representations of scientists, but also the content of scientific knowledge.

Scientific theories must be related to the genetically determined basis of human cognition in ways that make them 'cognisable' for humans. The relation involves complex causal chains that relate scientific knowledge to scientists' behaviour, and scientists' behaviour to their human biology (as well as to their personal history, esp. the social context in which they lived). How can we trace back the involvement of the universal structure of the mind in science? How is this involvement expressed in the content of scientific theories? One can inquire about the cognition of particular scientists, such as the particular analogies on which they draw for their discoveries. Some particular cognitive events do have an impact on culture. But the structure of the mind is also manifest in scientific stabilised practice. Atran (1990), for instance, has shown that intuitive thinking about natural species is still used, and useful, in day-to-day scientific practice, even though it is at odd with neo-Darwinist theories. The causal action of the structure of the mind on the content of science can also be made apparent by inquiring about the similarities or isomorphism between naïve and scientific theories: for instance, in my case study (chap. 7), I analyse the significance of similarities between the mathematical notion of the real line numbers and the mental representational system for magnitudes. I argue that the mental representational system has had a role in the history of mathematics, as it has constrained mathematical cognition on quantitites.

5.3 THE ODD COUPLE: NATIVISM AND SOCIAL CONSTRUCTIVISM

5.3.1 COMPATIBILITY BETWEEN NATIVISM AND SOCIAL CONSTRUCTIVISM

Mallon and Stich, in an article titled 'The Odd Couple' (2000) also note the compatibility and complementarities of the universalist approach to cognition of evolutionary psychology, on the one hand, and social constructivism, on the other. The universalist approach corresponds to what was described in the previous chapter (chap. 4) as cognitive descriptive monism (still drawing on Stich's vocabulary), and the social constructivists are those who argue that moral and epistemic norms are the product of social processes. Rather than pointing out the *ad hoc* assumptions enforcing incompatibility made by *some* social constructivists, Mallon and Stich trace the origin of the dispute to an implicit disagreement over theories of reference. Disagreement does not stem from different views about how the world is, but it stems from systematically giving different references to the same words. With the example of emotions, they argue that social constructivists refer, with emotion terms, to what is being described by the people studied themselves; the reference thus depends on the local 'ethno-psychology' that sustains people's description. Evolutionary psychologists, on the other hand, intend to use emotion terms with a scientific meaning independent of local folk psychologies. They then refer to universal psychological processes. It is possible to acknowledge the existence of both the universal psychological processes and their local implementations, on the one hand, and ethno-psychologies and its associated understandings of emotions, on the other.

The analysis can, as suggested by Mallon and Stich, be extended to other domains, such as rational thinking. So, while universalists talk about rationality as referring to some properties characterising human cognition, i.e. notions pertaining to scientific psychology, social constructivists refer to the local beliefs and norms constituting people's description of what is good thinking. Now, one can assume that social constructivists would not normally deny that humans are endowed with a capacity to think and that this capacity has some general properties to be uncovered by psychologists. Reciprocally, universalists cannot deny the existence of documented local beliefs about what is good thinking. This is the argument developed in the previous chapter (chap. 4) about actual cognitive processes as distinct from beliefs about what good thinking is. However, it is important to note that social constructivists would insist that people's understanding of emotions is *constitutive* of their emotions: it is not just that people from one culture to another carve the world at different joints; it is also that this very act constructs new entities. Thus, the local categorisations of emotions construct these emotions as part of the local people's mind. With a blank-slate view of the mind individuals' emotions can be seen as being imprint of cultural emotions.

The peaceful image of Mallon and Stich where evolutionary psychologists and social constructivists were telling different part of one single story fades away, because social constructivists can have more radical claims that the one they mention. With emotions, the empirical question of the plasticity of the mind reappears: social constructivists bet on the plasticity hypothesis, evolutionary psychologists bet on the existence of a fixed innate structure of the mind that they endeavour to discover. A similar tension comes out in the case of thought processes: social constructivists will hold that these thoughts processes are constituted by the cultural, normative environment of the individuals, while evolutionary psychologists will hold that they are importantly determined by hereditary, genetic, factors (genes working in interaction, of course, with the environment)

Nonetheless, there is still the possibility of forming a (fertile!) 'odd couple'. But this requires carefully choosing the Social Constructivist groom. Social constructivist assertions indeed can be assessed only by taking into consideration what is claimed to be socially constructed (Hacking, 1999). Social phenomena do certainly participate to the construction of things such as borders or national identities. Discourses about what it takes to be French or how to be a true Frenchman, for instance, are social events that participate to the construction of a French identity. Money is likewise socially constructed: it is through the social practices of exchange between goods and money that money is constructed as a means of exchange. A key mechanism is at work in these processes: beliefs and practices are constructing the object of which they are about. Beliefs about national identity construct the identity, beliefs about money (that it is valuable, for instance) turn bits of metal into money. The question, with Mallon and Stich's example, is whether cultural beliefs about emotions participate to the construction of these same emotions. This is an empirical question.

5.3.2 Evolutionary psychology and the Strong Programme

Because social studies of science currently never take into consideration the possible impact of innate psychological factors, it is easy to conclude that these studies assume the social construction of cognitive processes. After all, most of their works consists in showing the social determination in belief formation, and such social determination is much more easily obtained if the thinking processes are fully determined by the cultural environment. Yet, this assumption is not necessary for social studies of science to be worth the name of social constructivist. The social constructivism of the Strong Programme is essentially about the social construction of the conceptual category of science. Science, strong programmists would assert, is partly constructed by the beliefs that some given cognitive practices and some sets of beliefs are scientific (epistemic evaluation of beliefs, see sect. 4.2.3). Scientific knowledge, say sociologists of scientific knowledge, is what is so labelled by the scientific community. Taking what the community says about what is scientific and what is not as a first approach for the study of science is methodologically justified because it provides an empirical criterion for the demarcation of the object of study, thus fostering sociological research. It puts at the centre stage of naturalised studies of science the investigation of the processes through which consensus is achieved. Scientific knowledge is necessarily a knowledge agreed upon and whose scientific status is acknowledged. From this characterisation of scientific knowledge, it follows that it is socially constructed: the consensus of the scientific community upon what is to be qualified as science is achieved through social interactions.

At first glance, we are now in the case where cognitive scientists of the evolutionary psychology brand and social constructivists à la Strong Programme are telling different parts of a single story: cognitive psychologists explain which cognitive abilities and processes are involved in the formation of scientists' beliefs and sociologists describe the ensuing social processes: the communication of scientific results, the institutional processes through which scientists' assertion are assessed and the like. One will recognise in this account, however, a weak programme in the sociology of scientific knowledge, for the social processes above mentioned will not have a causal impact on the content of science. The Strong Programme claims that the social processes of assessment are socially implemented, but also that the normative criteria applied in the assessment processes are socially constructed. It is not only the modes of production and communication of scientific knowledge that is socially determined; it is also the content of scientific beliefs. In order to recover this causal impact, one needs to get back to the social factors at work in individual scientists' cognitive processes of belief formation. This is done in three steps:

1. The scientific norms and institutions as well as scientific theories and practices are cultural phenomena generated by individuals' behaviour (methodological individualism);

2. individuals' behaviour is largely determined by individuals' beliefs (relevance of psychological facts);

3. the scientific beliefs of individual scientists are constructed on the basis of the scientists' cultural background knowledge and social situation. (social determination of interpretation).

Starting from social phenomena, the explanation eventually ends up with the problem of the determination of belief formation. We thus come again to the 'sociologistic' assertions of the Strong Programme and its possible tensions with cognitive explanations (see sect. 3.3). What determines scientific belief formation, universal psychological factors or local interpretive traditions? The answer is both: local interpretive traditions are themselves understood and put at work through psychological processes enabled and constrained by properties that are shared by all (normal) human minds. Can psychological factors be reduced to social factors, because they are fully determined by the cultural context? The answer is no: some psychological factors come from the innate cognitive endowment of humans.

In order to sustain their claim that social factors enter the processes of scientists' belief formation, members of the Strong Programme have shown, in their case studies, that interests and social pragmatic considerations have a role in scientific belief formation. These social determinations are made most apparent in controversies, where the different social situations of the opposed parties can be shown to have a role through differentiation or co-variation of the beliefs. In these studies, the authors empirically trace the contextual factors having a causal role in belief formation and the consequent adoption or rejection of truth claims by the scientific community. The co-variation between scientists' scientific beliefs and their social situation, and the reconstruction of the influence of social context on reasoning (e.g. ideas about what society is or should be as a source of metaphorical thinking for understanding physical phenomena) are empirical evidence showing that the history of science is under-determined by the structure of the human mind and by human biology in general: contextual factors need to enter the explanation. Studying variation and, when possible, co-variation has been the most powerful methodological tool for pointing out the historical and social contingency of some scientific beliefs. Do we have similar methods for pointing out the dependency of scientific beliefs on psychological factors independent from the socio-cultural context? One important thing to do, in any case, is to be aware of the work done in psychology about these factors; interpreting the micro-social and psychological events of scientific thinking and communicating can then be done in the light of the results of cognitive psychology In particular, the analysis of scientific thinking must be done with the constraints set by evolutionary psychology.

5.3.3 Latour against evolutionary psychology

Not all school of schools in science studies would welcome information from evolutionary psychology. In the first chapter of his book Pandora's Hope (1999a), Latour considers evolutionary psychology as an alternative to the erroneous Cartesian picture of a 'mind-in-a-vat' observing the 'outside world'. He says: "Why not let the outside world invade the scene, break the glassware, spill the bubbling liquid, and turn the mind into a brain, into a neuronal machine sitting inside a Darwinian animal struggling for its life?" (p. 9). This image of evolutionary psychology rightly emphasize two themes: the biological, and thus material, basis of psychological abilities, and the importance accorded to biological evolution. However, Latour deems evolutionary psychology inappropriate. This is because, he says:

> the ingredients that make up this "nature", this hegemonic and all-encompassing nature, which would now include the human species, are the very same one that have constituted the spectacle of a world viewed from inside a brain-in-a-vat. Inhuman, reductionist, causal, lawlike, certain, objective, cold, unanimous, absolute — all these expressions do not pertain to nature as such, but to nature viewed through the deforming prism of the glass vessel! [...] Studying humans as "natural phenomena" [...] would abandon the rich and controversial human history of science — and for what? The averaged-out new orthodoxy of a few neurophilosophers? A blind Darwinian process that would limit the mind's activity to a struggle for survival to "fit" with a reality whose true reality would escape us forever? No, no, we can surely do better [...] retaining both the history of humans' involvement in the making of scientific fact and the sciences' involvement in the making of human history (p. 10)

There are two parts in this allegation against evolutionary psychology. The first part concerns the deforming prism of the glass vessel; the second part concerns the history of science. The second part is simply false: studying humans as natural phenomena does not mean at all abandoning "the rich and controversial history of science". Latour un-argued assertion seems to be based on the erroneous culture-nature dichotomy, which is at the basis of much of the visceral reactions to evolutionary psychology: the false idea is that if it is determined by nature, then it is not determined by historical factors. The proper argument for evolutionary psychology, however, is that human history is enabled and constraints by human biology — and in particular the biology of the brain. In particular, the mind's activity is not "limited to a struggle for survival", but it is produced by a biological apparatus that is itself the product of an evolutionary history. It seems obvious to me that the naturalisation of epistemology implies "studying humans as "natural phenomena""; and evolutionary psychology is, indeed, doing just that.

The first part takes the form of an obscure attack of the "ingredients" that make up the "nature" posited by evolutionary psychology. This attack is to be understood as included in Latour's constant attempts to discredit the discipline of psychology, which, according to Latour, has for only goal to maintain the myth of the mind-in-a-vat disconnected from a world totally outside, but still striving for absolute truth (Latour, 1999a, p. 13). Does evolutionary psychology persists in adopting the erroneous 'mind-in-a-vat' view of cognition, as Latour seems to claim? Yes, to the extent that it still consider notions as 'intention,' 'representation,' 'reference,' and 'belief about' as worthwhile topic of investigation. When Latour wants to do without such notions (see sect. 3.2.3, evolutionary psychology and teleo-semantics provide the tools and motives for naturalising these notions (see the work of Dredske and Millikan). So the more correct answer is that evolutionary psychology does not persist in the Cartesian view of the mind. Not only does it insist on the natural realisation of the mind, but it has also initiated and motivated an important trend of thought that insist on the situatedness of cognition, developing, in particular, the notion of 'ecological rationality' (Gigerenzer et al. 1999, the idea of ecological rationality is presented in section 5.4.3 below).

5.4 THE EPISTEMIC VALUE OF INNATE FACTORS OF COGNITION

5.4.1 PSYCHOLOGICAL RELIABILISM *versus* THE EVOLUTION OF CREDIBILITY

Evolutionary psychology, epistemic optimism, and reliabilist epistemology

Principles about the evolution of human cognition have been used in arguments about the epistemic value of science. One trend of arguments is that Darwinism shows that most of our beliefs are true, because natural selection could not have endowed us with deceiving cognitive abilities, since otherwise we could not have survived and made a successful species. If we have evolved a cognitive apparatus, then this cognitive apparatus provide us with true, rather than false, beliefs. The mental simulation is easy and telling: imagine that we were to have mostly false beliefs about, e.g. what is eatable and what is not, what is dangerous and what is not, the means to escape to a predator, the effects of falling down a high cliff, etc. If we were to have such false beliefs, then we would not have a great fitness. We would die rather quickly and with no descendant. The human species would not have survived. The brain is an organ that issue beliefs on the basis of which organisms act. Adapted action cannot but be based on true beliefs. So the brain issue mostly true beliefs. Truth-preserving inferential devices have been selected for during evolution and the world has imprinted on the human mind true innate knowledge.

Epistemic optimism about innate factors of cognition is, according to Stich (1990), "scattered here and there", but not developed in a full argument (I think it is still the case). Stich (p. 55) quotes Quine, Dennett, and Fodor, who all assert that selected organism will have true beliefs and will implement reliable inferences. Lately, Roger Shepard has more systematically defended the view that principles of the minds "reflect" principles governing the universe, and that organism with cognitive abilities have "internalised" biologically relevant (true) information about the world.

Epistemic optimism about our innate cognitive abilities can be combined with reliabilist epistemology, according to which reliable cognitive processes are those processes that mostly generate true beliefs, and knowledge is true beliefs acquired through reliable processes (Goldman, 1986). The argument has then the following form: evolution favour cognitive processes that are more reliable, i.e. processes issue more true beliefs than other, less reliable, processes, so genetically determined cognitive processes are knowledge providers. Innate factors of cognition offer sufficient guarantees that scientists will think properly. This is psychological reliabilism. It is the idea that rationality, *as a guaranty of truth* rather than as an empirical notion, is indeed implemented/realised in the mind. It is a foundationalist theory based on evolutionary argument.

SP and the evolution of credibility

Not surprisingly, this foundationalist theory has been used as an argument against the research programme of the sociology of scientific knowledge. Newton-Smith (1981) asserts that evolution shows that we are endowed with a natural rationality, to which one can appeal to distinguish reasonable from un-reasonable beliefs. Bloor rejoins Newton-Smith on his notion of natural rationality, but argues that appeal to natural cognitive propensities do not enable to decide which party in a scientific controversy is following the "dictates of reason" (Newton-Smith's term). The study of natural rationality is necessary to any comprehensive account of knowledge, but it is not sufficient for explaining why people endowed with natural rationality often disagree as to the 'correct' interpretation of the data; it is not sufficient for explaining what data is interpreted as evidence and for which theory (Bloor 1976, afterword of the second edition; for a detailed account of the dispute, see Kim 1994).

We are again in front of a case where epistemological beliefs about the mind preclude 'strong' sociological investigation of knowledge production. These beliefs go against the idea, maintained by sociologists of the Strong Programme, that what are reliable processes for scientific knowledge production is decided through social historical processes. Sociologists of the strong programme have argued that the credibility of scientific claims is dependant on the context. In particular, what is credible, acceptable as being scientific, has changed during the history of science. There has been an evolution of the 'evolvability' of scientific beliefs. This evolution of evolvability is at the heart of the sociologism of the Strong Programme: socio-cultural phenomena enter in the account of what is credible, because what is credible depend on the socio-cultural context.

The incompatibility is therefore not between the 'strong' sociology of knowledge and evolutionary psychology, but between the 'strong' sociology of knowledge and a reliabilist philosophy that would assert that all the reliability needed can be found in innate psychological factors. The Strong Programme, by contrast, asserts that which processes are reliable is decided in context.

I will argue that psychological reliabilism is false. It cannot be grounded in evolutionary considerations, because its basis — epistemic optimism about innate factors of cognition — cannot be extended to truth oriented scientific thinking. On the contrary, evolutionary psychology shows that there is a gap between the innate fitness enhancing properties of the human cognitive apparatus and the rational norms that sustain scientific practice. The naturalistic question about innate factors of cognition is not so much whether they are satisfactory for doing science as to how they make scientific thinking and scientific evolution possible.

5.4.2 EVOLVED COGNITIVE ABILITIES ARE NOT SUFFICIENTLY RELIABLE FOR SCIENCE

There is, in fact, no reason to think that our cognitive abilities are suited or satisfactory for the scientific enterprise. Stich argues against what I called psychological reliabilism by pointing out that, according to the principles of evolutionary psychology, our cognitive apparatus may not deliver 'mostly true beliefs' (1990, chapter 3).

The first reason why our cognitive apparatus may not be optimally reliable, i.e. deliver mostly true beliefs, is that evolution does not straightforwardly target optimal trait in order to implement them. Evolutionary theory, even of the adaptationist school, does not assert that if a trait is optimal, then it should exist. So granting that having true innate beliefs and truth preserving inferential devices is fitness-enhancing does not implicate that the human species is endowed with a cognitive apparatus with such properties. In particular, genetic mutations may never have put organisms with some really reliable cognitive apparatus to the test of evolutionary selection (see also Plantinga 1993).

The second reason why we do not necessarily have reliable minds is that more reliability, expressed in terms of the ratio true/false beliefs produced, does not necessarily imply more fitness. This is because fitness importantly decreases when effort, time and energy required by a cognitive organ increase. If the ratio true output/effort is good, then the cognitive device is efficient. Efficiency is a fitness-enhancing property since it leads to economize on rare resources. A dramatic example is furnished by the detection of predator: it is better when this detection does not take too much time; lest one end up eaten before reaching the conclusion. In such cases, a quicker process that is less reliable might be more fitness-enhancing than a slower but more reliable process. Efficiency can be favoured by natural selection even when it implies less reliability. Also, all errors do not have the same impact on survival. For instance, detecting a predator when there are none, is not a lethal error, while not detecting a predator when there is actually one, may be lethal. More generally, a cognitive system may contribute more to the fitness of an organism when it makes many errors with little consequence on survival and near to no lethal errors than a cognitive system that makes fewer errors on the whole but more lethal errors. The property of reliability, since it deals with the proportion of true and false beliefs independently of their content and consequences, is blind to such aspects of cognition, which are nonetheless strongly related to fitness-enhancement.

The general idea of this paragraph is that optimal cognitive processes are not necessarily maximally reliable, or maximally informative. Adaptanionist arguments for optimality should therefore not be understood as arguments for reliability. Fodor (1998, chap. 16) criticises both Plantinga

(1993) and Dennett (1978) for the epistemological conclusions they draw from evolutionary psychology. Fodor points out that the evolution of our minds furnishes no warrant about the truth of our beliefs. Contrary to Dennett's argument, Darwin does not help to find an epistemological foundation of science. But contrary to Plantinga's argument, Darwinism does not imply either that the sceptic is right. Fodor reminds us that we believe the truth of scientific theories because of the evidence that are advanced in favour of them, rather than because of the fact that human cognition evolved (let me add, with the Strong Programme, that what is taken as being an evidence is context dependant!). Hull (2001) arrives to the same conclusion: "adaptation does not guarantee truth" (p. 162). As one of the main protagonist of evolutionary epistemology, he asserts, "the fault with evolutionary epistemology lies not with it being evolutionary but with its being epistemology" (p. 163). This is to say that evolutionary theory has no obvious consequence on how we can evaluate science. This opens up the possibility of elaborating consequential means and criteria for the production of truth, which is what scientists and philosophers have been doing during the history of science. There is a history of epistemological beliefs, there is an history of means and criteria of credibility.

5.4.3 Dissatisfaction with cognitive performances

Another argument against psychological reliabilism comes from the observation of how people actually reason: *de facto*, people reasoning processes are, in many situations, not reliable. This was the striking result of a large set of psychological studies well developed since the 80' about cognitive biases. While these experimental results are normally used for what they reveal about human individual rationality, I want to point out the discrepancy between initial responses, directly driven by innate dispositions, and the criteria through which we eventually evaluate the results. The conclusion I derive from this discrepancy is that the evaluative criteria are not the direct image of our cognitive abilities: they need to be elaborated. The elaboration of the evaluative criteria is then a new object of inquiry. It is

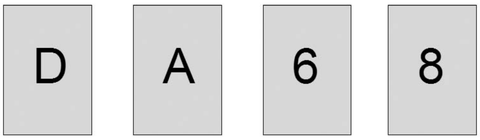

Figure 5.2: The Wason task

the existence of this object upon which the Strong Programme has insisted, in order to justify its research programme.

Cognitive biases

Among the most famous experiment showing that individual do not necessarily comply with truth preserving principles of rationality is the Wason task (or selection task), which shows that subjects do very poorly at understanding the implications of a conditional. The usual experiment is presented in the graph 5.2:

Here are four cards. Each has a letter on one side and a number on the other side. Two of these cards are with the letter side up, and two with the number side up: Indicate which of these cards you need to turn over in order to judge whether the following rule is true: **if there is a D is on one side, there is a 8 is on the other side**

In this form, only very few subjects (around 10 %) find the correct answer (the card marked with D and the card marked with a 6).

Another of the most studied deviation from normative standards of inference is the conjunction fallacy, where subjects find a conjunction (A and B) more probable than one of its conjunct (say, A). Tversky and Kahneman's (1983) experiment is as follow:

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Rank the following statements from most probable to least probable:

1. Linda is a teacher in an elementary school

2. Linda works in a bookstore and takes Yoga classes
3. Linda is active in the feminist movement
4. Linda is a psychiatric social worker
5. Linda is a member of the League of Women Voters
6. Linda is a bank teller.
7. Linda is an insurance salesperson.
8. Linda is a bank teller and is active in the feminist movement.

In this experiment, 89 percent of the subjects erroneously ranked (viii) as more probable than (vi).

Other experiments have shown that subject systematically (and erroneously) ignore base rates in probalistic reasoning, that they are very reluctant to change their prior beliefs in the face of evidence and that they do not recognize valid inferences as valid if the conclusion seems improbable to them.

Ecological rationality

Much work in psychology has been devoted to interpret these results, which went against the a priori beliefs that humans are rational in the sense that they reason in accord with acknowledged norms of reasoning. One conclusion is that humans often rely on heuristics when taking decisions. Heuristics are decision procedures that dispense with rational analysis and output judgements on the basis of few informative cues. Bypassing rational analysis means doing without the best guarantee of the truth of the output, but it also has for consequence a great economy of computational time and effort. Indeed, in most natural cases, rational analysis is not even computationally possible, because it implies reviewing too much information. The problem is not solved when rational analysis is understood as maximising cognitive time and energy — since they are rare resources — for a dilemma arise regarding information gathering: how to know whether it is worth gathering further information, since we cannot know the worth of the information not yet gathered in advance? Gigerenzer and his collegues (Gigerenzer et al., 1999; Gigerenzer and Selten, 2001) have argued that another notion of rationality had to be developed in order to account for humans' cognitive processes. The fact that people do not comply with the psychologist's norms of good thinking does not make them dumb or simply irrational. Rather, human cognition is best characterised as being ecologically rational. Ecological rationality is a property of cognitive processes that answer the fitness-enhancing requirement of natural selection much better than the traditional notion or rationality. Indeed, while the traditional notion is focused on truth-preserving properties of inferences, the notion of 'ecological rationality' takes into account the goals and means of the organism, as well as the specificity of its environment. Gigerenzer and his colleagues have therefore shown that "fast and frugal" heuristics do better in given contexts than more complex cognitive processes that satisfy better the traditional criteria of rationality. Heuristics do especially well when they are adapted to the environment. For instance, frogs throw their tongs at small black flying objects, with the consequence of feeding themselves with flies. The cognitive process is therefore meant at catching flies, but it relies on few cues only: being small, black and flying. The cognitive process is not based on a general truth, since not all small black flying things are flies; and it is neither error free, since frogs can be in the situation of swallowing non-nutritive elements. Yet, the heuristic 'throw your tong at small, black, flying things' is nonetheless ecologically rational because, in the environment of the frog most small black flying things are flies indeed. The success of the cognitive process is based on a statistical property of the environment, rather than on fine-grained analysis.

perceptual illusions

Dissatisfaction with our cognitive abilities is not restricted to difficult cognitive tasks. It extends to perception, as is shown with the much studied

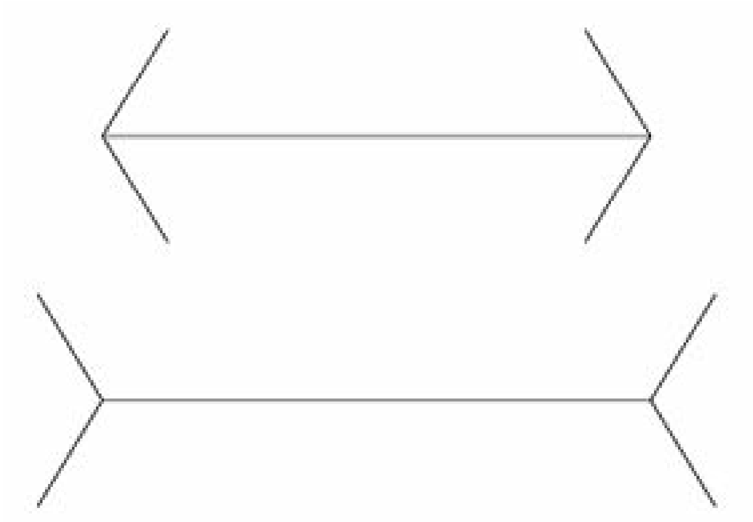

Figure 5.3: The Muller-Lyer illusion: while the top line (arrows pointing outside) appear shorter than the bottom one, the lines are in fact of the same length.

visual illusions. The usual example is given by the Muller-Lyer illusion (5.3.

Our visual apparatus, together with brain processes, allow us to have a fairly good intuition of distance. We are able to judge if something is far or near, big or small and if something is farther, bigger or longer than something else. This ability therefore amount to a biologically embodied knowledge of distance that has been selected during the course of evolution. Yet, although our ability is reliable and guide behaviour so as to adjust the trajectory of our body in a crowded environment or even throw things at targets in a rather successful way, it does not provide a scientifically satisfactory criterion.

I have presented two arguments: fitness-enhancing cognitive processes are not necessarily reliable processes, and humans are prone to systematically deviate from our own criteria of rationality. The conclusion of these arguments is not that thinking processes always lead to false beliefs, and it is admitted that true beliefs are, in general and *ceteris paribus*, more fitness-enhancing than false beliefs. One conclusion is that evolutionary psychology does not furnish the guarantee that our cognitive abilities are reliable. The second conclusion is that there is an epistemic evaluation of the output of our cognitive abilities.

5.4.4 Is there any scientific-friendly evolved cognitive ability?

Evolutionary psychology shows that the biological function of the human cognitive apparatus should not be confused with the explicit goal of science (contrary to what is implied by Gopnik 1996). It is only the scientific entreprise that look for truths for the sake of it, and with a relative independence from pragmatic consequences. The aim of our innate cognitive endowment is not to provide the best knowledge on the basis of the available evidence. This cognitive endowment has rather evolved to guide behaviour so as to improve the chance of survival and reproduction. This situation could be interpreted as giving some grounds to the sceptic, but the naturalistic reaction is rather to question how scientific is made possible in this conditions. There must be some scientific-friendly evolved cognitive ability.

Psychological reliabilism can accept that some cognitive processes are not sufficiently reliable for scientific investigation but that we have nonetheless some other abilities that are sufficiently reliable. The obvious evidence for such a view is that we do manage to think scientifically in spite of natural biases impeding good reasoning. Some psychologists have hypothesised that we have an evolved ability for proper reasoning, which they call *system 2* (see esp. Evans 2003, who appeals to Mithen 2002 as showing the evolutionary plausibility of his psychological hypothesis).

There are, however, theoretical problems concerning the evolutionary history of a domain general cognitive device that would implement truth preserving cognitive processes: on the basis of which antecedent could this device have evolved? What is the adaptative value of such a device? Ecologically rational cognitive devices will always have the advantage over a domain general system 2 device, because they will be more efficient in dealing with specific types of problems. If system 2 reasoning is concerned with scientific reasoning, then it is certainly not an evolved cognitive device with some genetic basis, since science appeared only lately and locally in the history of humankind.

Another strategy for understanding the possibility of scientific thinking and reasoning is not to assume that it is a unitary phenomenon that is well distinguishable. There are no obvious demarcation criterion between science and non-science; and there is no demarcation criterion between scientific thinking and non-scientific thinking. Scientific cognition is enabled not only by our innately specified cognitive abilities, but also, and essentially, with the cultural context with which the human cognitive apparatus interact. There can be variations in ideas about what good reasoning is, and the cognitive event to explain is that individuals are able to conform to these variable ideas. This suggests that we have a capacity to develop and comply with variable norms of good reasoning. This sends us to the human interpretive and evaluative abilities that I described in the previous chapter (4.2.2). The cognitive biases are understood as biases because evaluative, meta-representational abilities are put to work for evaluating our own initial responses.

frame effect as the experimental counterpart of the history of credibility

One striking result of the studies of experimental psychology on cognitive biases is that the biases can disappear when the question is framed differently. The structure of the problem remains the same, but the context is changed. One plausible reason is that changes in the frame change what information is salient and relevant, and ideas about what is expected in the social context of the psychological experiment. I submit that much of scientific evolution consist in framing scientific problems. The problem of incommensurability appears not necessarily because scientists, in a scientific controversy, have different and incommensurable scientific concepts and carve the world in different ways. The problem of incommensurability can be understood as the fact that scientists elaborate their own frames for thinking about similar problems: the structure of information, what is salient and relevant, is not necessarily the same. Works in the sociology of scientific knowledge suggest that much of the so called negotiations are negotiation about how to frame the problems and what information to take as more relevant.

A similar process is put in place for perceiving visual illusion as illusory. In order to realise that the lines of the Muller-Lyer illusion (5.3) are of the same length, we can rely on a cultural practice, which consist in comparing the length of lines by laying one onto the other. So you may fold the paper and see that they are in fact of the same length... or you may also appeal to a more complex practice: you can use a ruler, lay it onto the first line, remember the number where it stopped, lay it onto the second line and compare the number obtained. Another solution is to hide the arrows at each extremity. Appealing to these practices is not throwing away all intuitions: the practices still rely on intuitions, for instance, when after juxtaposition we see that the lines overlap, and we also see next to which number the line ends on the superimposed ruler. Yet, the practice is a socio-historical construct with a high normative impact. In order to correctly compare length, one must correctly go through the steps prescribed by the practice. There are frames where our cognitive apparatus issue true beliefs and other frames where it is misleading. In order not to be mislead, one must pick up the *correct* frame.

Being social as being rational: argumentative abilities

In the previous chapter (chap. 4), I suggested that social cognitive abilities and naïve psychology were at the heart of our ability to reason in accord with norms of good reasoning. One necessary condition for complying with norms of reasoning — seen as social construct — is to understand others' expectations about how we should reason. Reciprocally, these expectations are realistic: they are based on actual performances and the naïve psychology of the expecting person, with naïve psychology being good enough to adequately foresee others' behaviour. Sperber (2000, 2001c) also argues that the social environment in which humans have evolved has rendered possible and likely the evolution of abilities that assess the validity of what others communicate. The important specificity of these abilities is that they are both ecologically rational, in the sense that they are adapted to the socio-cultural environment of humans and have a plausible evolutionary history, and oriented towards epistemic validity. Thus, Sperber hypothesises the existence of a "logico-rethorical" module, which constitutes an "ability to attend to formal, and in particular to logical properties of representations." In spite of the name, Sperber does not hold that the logico-rethorical module is the psychological mirroring side of mathematical logic. He also insists that this ability is not domain general but deals only with abstract properties of representations — and more specifically of communicated representations. The logico-rethorical module answers a selective pressure that arise in highly communicating species and which strongly privilege those organisms that can filter communicated information. Spotting out deceiving information on the basis of its logical structure provides a decisive advantage and might well be a requisite if communicative abilities are to evolve: indeed, without this filter, communicating deceiving information is fitness-enhancing for the deceiver (you get people doing what you want them to do), so deceiver should spread, causing in the long term the disappearance of organisms to deceive (they cannot compete with deceiver for food, mates, etc.) and the uselessness of communicating abilities. The logico-rethorical module is thus conceived as an ability that protects against misinformation, which then triggers the evolution of argumentative abilities as means to convince those who are endowed with such filtering abilities. We have the ability to do evaluate the truth or probability of communicated information and, more generally, to be epistemically vigilant; how can we explain this ability? The fact that communicative ability have evolved suggests that semantic evaluations are based on evolved abilities or modules dedicated to filter which communicated information should be believed and which empower communicative abilities with rhetorical means. Now, if one admits that communication is an essential aspect of science, then the evolution of science will essentially rely on these logico-rhetorical abilities. At the outset, this is no great news: everybody knows that scientists communicate and that part of their job is to check the validity of each others' statements. Yet, this displace much of the so called logical or rational aspects of scientific practices from the level of the individual investigator thinking about the world, to the level of communication of information, where thinking about the world include thinking about one's collaborators — what they already know, what they say, the questions they've raised, and how they could be convinced. In other words, scientific thinking makes most use of abilities that correspond best to the classical idea of rationality at the social stage. Beller (1999) has shown, with a historical study in the history of quantum mechanics, that scientific production is essentially a dialogue between scientists. Scientists never cease to address each others. I submit that inner dialogues are constitutive of scientific thinking and pervasive in scientific cognition. A key aspect of inner dialogues for scientific cognition is that it taps in rhetorico-logical abilities, thus activating early on in the process of scientific knowledge production such processes as partial consistency checking between sets of relevant beliefs or other processes pertaining to our innate abilities of 'epistemic vigilance.'

How, then, do we obey the scientific principles of reasoning, given that our minds is essentially ecologically rational? Part of the answer lay in our interpretative abilities, our meta-representational ability to evaluate the truth of representations, our communicative abilities, and the human abilities to be epistemically vigilant. All these abilities seem at the heart of scientific cognition, however, they do not *replace* other cognitive abilities — on the contrary, they establish ways to exploit those abilities through evaluating their output and framing the problems so as to activate their inferential power.

Conclusion not every social constructivism is compatible with evolutionary psychology And not every interpretation of the epistemological implications of evolutionary psychology is compatible with social constructivism. Yet, I have argued that we can have the couple evolutionary psychology – Strong Programme. A couple which is not so odd with all that, since Bloor and Barnes have done some preliminary courting. The wedding, as often, requires dealing with important questions and possible problems. One of the challenge that one needs to confront is the possibility of the historical variability that is observed by sociologists of science, given the heavy innate constrains on human cognition posited by evolutionary psychologists. This is the question I have tackled in the previous chapter. Another challenge is then to discover which aspects of 'natural rationality', or of our evolved cognitive abilities, have exerted a constraint in the historical production of some specified scientific theory; how these constraints operated and how they are manifested in the content of scientific knowledge. The ability to exploit interpretive traditions in scientific thinking frees scientists of much of the constraints posed by their limited cognitive abilities. This is why the history of science must be a social history about which interpretive traditions where developed, taken on and why. Does that mean that, eventually, the constraints set by the nature of the human mind are irrelevant for understanding scientific development? I have suggested that the answer is no. Not all interpretive traditions can be processed equally well by the human mind. Interpretive traditions must be appealing so that scientists use them, and this appeal arises from their relations with scientists' minds, which includes: how they are understood, how they are memorised, which inferences they enable to make when combined with new input, and which inferences they cause when combined with past knowledge The species specific aspects of humans significantly determine all these cognitive processes because they are done by specialised cognitive device with their own import in naïve knowledge and specific inferential mechanisms. The next chapter attempts to spell out the determining role of evolved cognitive abilities in the history of science. In this chapter, however, I have introduced notions from cognitive psychology, mostly evolutionary psychology, with which on can question anew the relations between scientific cognition and the nature of the human mind. The latter is characterised as ecologically rational and is endowed with a number of domain specific abilities. How are they put at work in the pursuit of scientific knowledge?

Chapter 6

The epidemiology of scientific representations

Dan Sperber has elaborated a theoretical framework — the epidemiology of representation — that enables taking into account the psychological factors explaining and shaping cultural phenomena. In this section, I first present the framework; second, I specify why and how it could be applied in science studies. In the third section of the chapter I attempt to describe characteristic features of scientific cognition and evolution.

6.1 The epidemiology of representations—a brief account

The epidemiology of representation aims at satisfying the two goals mentioned in the epigraph of the chapter: providing a naturalistic characterisation of culture, and furnishing a theoretical framework for explaining cultural phenomena that would take into account the importance and relevance of psychological factors.

6.1.1 A naturalistic characterisation of culture and its evolution

Providing a naturalistic characterisation of culture is, for Sperber, an ontological requirement that forbids being fully satisfied with macro-social explanations:

> From a naturalistic point of view, we must either dispense with

> such macro-entities, or unpack them in terms of micro-phenomena. To reconceptualise the field we may draw inspiration from a science that is at once social and natural, I mean medical epidemiology. In epidemiology, social macro-phenomena such as endemic and epidemic diseases are unpacked in terms of patterns of micro-phenomena of individual pathology and inter-individual transmission. (Sperber, 2001a)

Sperber's naturalisation is therefore an attempt to grasp social phenomena without appealing, in the explanation, to concepts that do not refer to natural entities — that is to say entities whose physical properties are not well understood, whose material existence is not specified, such as 'marriage', 'kinship system', 'myth'. Two natural entities are constitutive of cultural phenomena: mental representations and public productions. Public productions are "any kind of object in the environment that humans can produce and perceive"; they are bodily movement, utterances, written symbols, works of art, etc. Public productions include public representations, which are productions that are indented to generate mental representations in the people that perceive them. The material character of public production is relatively unproblematic. Mental representations are being naturalised through the research done in cognitive science (which investigates in particular how mental events may be implemented in the brain). With this minimal ontology, Sperber states: "widely distributed, long-lasting representations are what we are primarily referring to when we talk about culture" (Sperber, 1985, p. 57). More precisely, cultural representations are types of representations whose token representations are well distributed in the human population and its habitat; where types of representations are constituted of sets of mental or public representations that members of the cultural group judge sufficiently similar to one another so as to constitute a type (word tokens as being the same word, narratives as being of the same tale, food on their plate as being the same dish, performances as being the same ritual, etc.). The epidemiology of representation seeks to explain cultural phenomena in terms of the mechanisms through which representations stabilise in a population, so as to qualify as cultural. These mechanisms, Sperber argues, take the form of cultural cognitive causal chains, which are defined in three steps (Sperber, 2001a):

Cognitive Causal Chain A causal chain where each causal link instantiates a semantic relationship

Social Cognitive Causal Chain A cognitive causal chain that extends over several individuals

Cultural Cognitive Causal Chain A social cognitive causal chain that stabilises mental representations and public productions in a population and its environment

The chains are cognitive because causal links instantiate semantic relationships that best explain the link between the cause and the effect. Semantic relations best explain the causal relations when the mechanisms producing the causal process have the functions of securing semantic relationships, i.e. when the mechanisms are cognitive. A brain, for instance, can be considered as including a set of cognitive mechanisms. Thus, when I hear the bell ringing, to follow Sperber's example, I infer from it that someone wants the door open, which "is a causal process that takes as input the general representation that what normally causes doorbell to ring is the action of people wanting the door open and the specific representation that the doorbell is ringing." Further representations can enter the chains, as my remembering having ordered a pizza, and so on. Social cognitive causal chains include cognitive causal chains and social interactions, which implicate public productions as means to extend the cognitive causal chain from one individual to another. Two links are specified by Sperber, with the help of the graph 6.1 (taken from Sperber, 2006a):

Social cognitive causal chains can stabilise representations when they involve many people in time and space and produce representations with similar contents. The telling of a tale from parents to children across generations is an example that is made of a relatively simple cultural cogni-

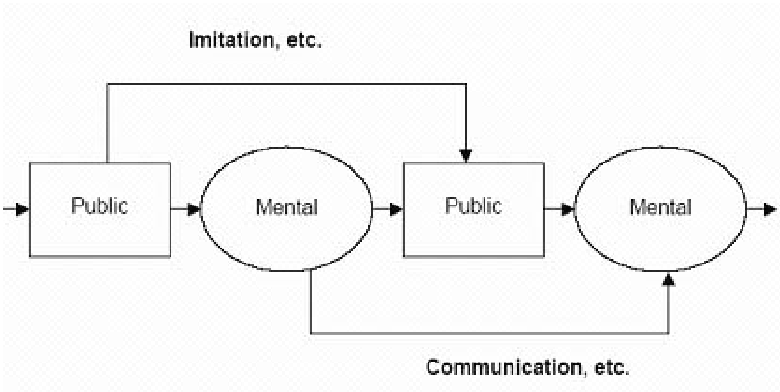

Figure 6.1: Two kinds of links in social chains

tive causal chain. Other chains may be more complex and involve a great many different representations related in different ways, such as the cultural cognitive chains that stabilise social institutions. French civil marriage, for instance (Sperber, 1996a, p. 30), involves representations about the ceremony, about the role of the civil officer, held by the people who perform the ritual. Public productions such as books of the *Code Civil* are also implicated in the chains. Utterances are uttered again across ceremonies with little variations, such as "I hereby declare X and Y husband and wife".

One can distinguish two factors that contribute to or hinder the stabilisation of representations: environmental factors and psychological factors. This corresponds to the fact that cultural cognitive causal chains are made of events that happen inside or outside the organisms' cognitive apparatus.

> The environment determines the survival and composition of the culture bearing population; it contains all the inputs to the cognitive systems of the members of the population; it determines when, where and by what medium transmission may occur; it imposes constraints on the formation and stability of different type of public productions. (Sperber, 1996a, p. 113)

Paths are public productions that are determined, among other things, by features of the landscape (flora, relief ...) because such features will influence the choices of the walkers. "The slope of roofs tends to be stabilized by local weather conditions" (2006a). An utterance lasts only the time when it is uttered while a book can last for generations; books can be read only on the place where they are and are not easily reproduced, while electronic web documents can be instantly reproduced on users' screens. The environmental factors, it should be noted, can themselves be culturally shaped: the technology of communication just mentioned is a case in point. "[M]uch of our cultural practices is stabilized by the affordances and constraints of cultural productions", concludes Sperber (2006a).

Cultural cognitive causal chains importantly implicate the working of the human cognitive apparatus, so psychological factors intervene in the mental processes that produce mental representations and determine behaviour. This leads to the second point of Sperber's theoretical framework: showing why and how psychology matters in explaining cultural phenomena (i.e. psychology not only helps characterising culture in general, but it also enable explaining aspects of specific cultural phenomena). Understanding the causal chains in which the mental representations and public productions are involved implicate understanding the causal events that happen in the head of the individuals engaged in the chains. Sperber strongly contrasts his views with other naturalistic accounts that conceptualise culture in terms of population of mental and artifactual items distributed in the human population and its environment, but which tend to over-simplify or misrepresent the cognitive processes implicated. Memetics (Dawkins, 1976) is such a theory of culture: it applies Darwinian selection for modelling cultural evolution, and erroneously assumes that cultural items are replicated through imitation. But representations are not simply replicated by the human brain, they are not simply copied from one mind to another; and the mind does not either average out similar representations received as input. Communication, in particular, is not a process through which the mental representations of the speaker are coded into a public representation, which is then decoded by the audience. What happens, rather, is that the audience is building representations about what the speaker intended to communicate when producing linguistic public representations. The cognitive processes involve the ability to attribute intentions to others, called in the psychological literature mind-reading abilities or theory of mind. What generally follows is that the processes of transmission are not wholly determined by the input to the receiving organism; they are also determined by how the input is processed. The epidemiology of representations therefore calls for a fine grained understanding of the parts of the cultural causal cognitive chains that are located within the heads of the people involved in the chains. It furnishes a framework that describes cultural phenomena in such a way that the relevance of psychology for the explanation of the cultural phenomena becomes apparent. Psychology is to the epidemiology of representation what pathology is to medical epidemiology.

6.1.2 Psychology and the stabilisation of representations

Psychological factors shape cultural phenomena because of two things: first mental processes have a pervasive role in the stabilisation of representations among a community; second, these mental processes shape the representations that are being stabilised. Representations are transformed across social cognitive causal chains rather than just being replicated, but they are not transformed in a random way. If it were so, representations would not stabilise and culture would not be possible. In order to understand how and why representations are transformed, one needs to peer into the non-cultural factors of though processes and, in particular, in the innate endowments of the human mind. Sperber draws upon nativist theories of the mind, evolutionary psychology and has been a main proponent of the massive modularity hypothesis (see section # cross ref 'chap. 4, 1.1. why the structure'). In addition, Sperber holds that human cognition is geared so as to process information for maximal effect and minimal mental effort, where effect is achieved when new information is made available for action or further thinking, and effort is done when processing representations (inferring, retrieving old information, etc.) (Sperber and Wilson, 1986, postface to the second edition). The effect-effort balance determines the degree of relevance, which qualifies input to cognitive devices: if the input has great cognitive effect and demands little effort when processed, then it is a relevant input for the cognitive device. Thus, the assertion of Sperber and Wilson is that the human brain tends to maximise relevance: it selects which available information to attend to, and which available past information to process it with. Communicated information in particular is presented by its producer as relevant to its audience and this presumption of relevance guides the audience in comprehension. Also, humans are social animals who are very good at manipulating each others (see the literature on Machiavellian intelligence - #REF). They are motivated to influence others' thoughts and they are rather good at it. In order to attract and direct the attention of others, humans tend to produce representations that are relevant to their audience. They thus increase the probability that their production will be attended to.

Relevance theory and the massive modularity hypothesis form the equipment with which Sperber exposes the psychological factors that shape culture. The causal action of psychological factors in cultural cognitive causal chains is based on at least the following psychological facts:

- input will be processed with the available memorized information and the specialised cognitive devices, the mental modules, of the individual;
- representations that are relevant will be more attended to than representations that are less relevant;
- public representations will be processed with the presumption that they are relevant;
- humans tend to produce representations that are relevant to their audience.

As a first example of how psychological factors shape culture, Sperber makes this mental experiment:

> suppose that an incompetent teller has the hunters extract Little Red Riding Hood from the Big Bad Wolf's belly, but forgets the grandmother [...] hearers whose knowledge of the story derives from this defective version are likely to consciously or unconsciously correct the story when they retell it, and, in their narrative, to bring the grandmother back to life too. In the logical space of possible version of a tale, some versions have a better form: that is, a form seen as being without either missing or superfluous parts, easier to remember and more attractive. (Sperber, 1996a, p. 108)

Why is the defective version of a bad form? Sperber does not tell at this point, leaving it to the intuition of the reader that the normal form is more attractive than the defective. A hypothesis, however, is that it is understood that the hunters are benevolent, since they are acting against the bad wolf. But benevolent people cannot so easily forget about the well being of a grand-mother — this is, at least, what our Theory of Mind leads us to expect. Of course, cultural information also plays a role in the fact that we want to satisfy this expectation: the tale is intended for children for whom maximal relevance may be attained when explaining how difficult situations can be sorted out (one just needs some benevolent hunters). Imagine however that the tale was to be told to teenagers among whom poking fun of grand-mothers is highly appreciated. Telling a tale that goes against the expected behaviour would increase the relevance of the story: it would raise the question 'why did the hunters leave the grand-mother in the wolf's belly?', opening up inferences such as 'leaving grand-mothers in wolves' belly is more benevolent than taking them out', 'we are better off when grand-mothers remain in wolves' belly' and so on. In both cases, both modular abilities (I mentioned the working of Theory of Mind in producing expectations about the hunters' behaviour) and cultural background information play a role as factors of attraction towards a specific form of the tale. Cultural phenomena are shaped by both psychological and environmental factors of attractions, which attract the production of representations, mental or public, towards a form where they achieve maximal relevance. Sperber calls this ideal form an 'attractor'. In the above example, the tale is likely to be remembered and produced in versions that are more than less similar to the attractor. The form or position of attractors depends on the cultural environment, since the cultural environment makes available information vary and since the relevance of an input depend on the information with which it can be processed. So Sperber talks about cultural attractors. Fashions, for instance, are quickly changing cultural attractors.

One way to achieve relevance is to exploit the inferential potential and the proprietary information of modules. This is achieved when the input triggers some modular abilities which consequently enrich its cognitive effect at little cost. As specialised cognitive devices, modules do not treat every input. They process only those inputs that meet some criteria on the basis of which it is likely that the activation of the module will be beneficial for the organism. For instance, frogs have a catch-flies module that triggers their throwing their tongues at small black moving things. The input conditions of the catch-flies module are: the entity is black, small and moving. In the environment of frogs, these input conditions enable them to throw their tongues at flies, and mostly at flies — although there can be small, black and moving things that are not flies. Input that meet the input criteria of a module are said to fall within the actual domain of the module. Such input directly 'benefits' (i.e. is processed with) from the database of this module, including innate knowledge, with which it can be processed; and it benefits from 'hardwired' quick and low-cost cognitive processes. For instance, face recognition is hypothesised to be modular in humans: we have a special ability to recognise faces and their expression which is incredibly effective with regard to the complexity of the task. Computers are very bad at recognising faces on the basis of an analysis of facial features, while we do it quasi-automatically and with little effort. The face recognition module constitutes a psychological factor of attraction towards the productions of artefacts that fall into the domain of facial recognition, viz. masks and portraits, or make-up (Sperber and

Hirschfeld, 2004). More generally, "a reliable way to attract attention is to produce information that falls within the actual domain of modules", explains Sperber and Hirschfeld (2004), so one can expect that some cultural representations will be found in the vicinity of the actual domains of modules. Sperber and Hirschfeld consequently characterise cultural domains of modules as the specific type of information that is culturally produced to activate a module. Across his work, Sperber mentions several cultural domains: totemism, mythic animals as cultural domains of folk biology; masks as a cultural domain of face recognition; music as a cultural domain of some early proto-speech recognition device; numeracy as a cultural domain of a cognitive device for thinking about quantities; cultural identity construction, such as caste or race, as a cultural domain of folk sociology — a competence which evolved for understanding basic social grouping. But mental modules' factors of attraction are not limited to the production of cultural domains of single modules. Religious beliefs are cases in point. The relevance of religious representations stems from limited violations of intuitive expectations. Intuitive expectations are produced by representations that fall within the domain of a module. For instance, talking about a person triggers the Theory Of Mind module which produces expectations about the attribute of the person — that the person can entertain a limited number of thoughts or that her knowledge is dependant on her experience for instance; talking about a physical entity also generates expectations about the behaviour of the physical entity: that it is subject to gravity and that it cannot go through a wall. These expectations are violated by supplementary explicit representations such as 'God can see everything' or 'Ghosts go through walls', which are attention grabbing, memorable and rich in inferential potential. In each case, the input provides important cognitive effects at low cost by exploiting the inferential power of some modules: this results in low cognitive effort — as in the case of faces, which are processed easily — and/or high cognitive effect — as in the case of counter-intuitive information, which opens up a new range of possibilities. In brief, psychological factors of attraction are based on the added relevance stemming from module's inferences, which are low cost and potentially numerous. Attraction is then around inputs that obey at least one of the following principles:

1. The input activates a module with little preliminary effort; it squarely fall within the actual domain of the module. The input is thus processed with low cognitive effort (e.g. smileys for communicating feelings reliably and at low costs; make-up as 'hyper-stimuli' for the face-recognition module).

2. The input activates a module and brings new information to it (i.e. not already figuring in its proprietary database), thus generating numerous inferences. The input thus issues high cognitive effect (e.g. counter-intuitive assertions in religions).

The means through which representations can stabilize in a population are not limited to the direct exploitation of a given mental module. In particular, Sperber insists on the role of deference in cultural transmission: some representations are stable because they are transmitted by people who have the authority to recommend to others for attention and acceptance transmitted representations. Institutions are also involve self-regulating, stable sets of representations (see #cross-ref). The epidemiology of representations thus resembles more a research programme than an exhaustive theory of the social. Its basis is a characterisation of cultural phenomena:

> Some information, being of more general relevance, is repeatedly transmitted in an explicit or implicit manner and can end up being shared by many or even most members of the group. 'Culture' refers to this widely distributed information, its representation in people's minds, and its expressions in their behaviors and interactions. (Sperber and Hirschfeld, 2004, p. 40)

This characterisation raises a specific question: how do mental representations and public productions stabilise in a population and its habitat?

Elements of an answer have been brought by Atran, Boyer, Hirschfeld and Sperber especially concerning the psychological factors of stabilisation.

Cognitive theories of social phenomena are dependant of the psychological theories the social scientist draws upon. The main trend in cognitive anthropological theories, for instance, has drawn on Rosch's work on concepts and Johnson-Laird's theory of mental models. Sperber and his colleagues draw on a still controversial, but more and more successful corpus of psychological literature which mainly asserts that there are conceptual modules — mental devices that deal with specific conceptual input. They have shown that these devices have an important role in shaping culture. The argument I develop in this part of the thesis is that they have a role in shaping one specific cultural phenomenon: scientific knowledge.

6.2 The epidemiology of representations and the history of science

6.2.1 Questioning the distribution of scientific representations

The epidemiology of representations provides a good theoretical framework for the history of scientific ideas. First, the epidemiological framework allows for a diachronic analysis of cultural phenomena, which are characterised as distributions of representations *in time*, as well as in space. Cultural representations are made of token representations that remain recognizably similar to *past, antecedent*, token representations; these token representations must span relatively long period of time in order to achieve a cultural status. The cause of the existence of cultural representations is also inscribed in time: Sperber points that the social cognitive causal chains that stabilise representations are "long and lasting" (2001a).

Second, questioning the processes through which representations stabilise is a good way to question the principles of scientific evolution. Studies in the epidemiology of representation (i.e. the work of Atran, Boyer, Hirschfeld and Sperber) have focused on the conditions, and in particular the cognitive conditions, which allow for the resilience and the continuity of content of some representations. This focus is justified because most social events do not lead to stable representations. Therefore, something more is needed to make a representation cultural, and it is this something more that is important for explaining culture. The question of stability is, for the same reasons, important in the history of science: most representations produced by scientists do not stabilise. Why and how do scientific representations happen to be successful? Why some are taken on and spread, while others are rejected or simply forgotten? The central question of the history of science can therefore be rephrased in epidemiological terms: what are the processes of distribution of representations among the scientific community that enable such and such representations to stabilise and such and such other representations to cease to be important and widespread?

Yet, what has interested most the historians of science is the development of science — not so much how some beliefs last, but mostly how and why beliefs change. Cultural change, however, also falls into the epidemiological rationale: the stability of a cultural representation is always partial and depends on the environmental and culturally contingent conditions that sustain it. Studying how changes in the environmental conditions affect the stability of a representation is indeed a genuine epidemiological question, i.e., one that can fruitfully be answered within this theoretical framework.

Another aspect of the epidemiology of representations that makes it an appropriate framework for the history of science is its stress on individuals' cognition as constitutive of cultural phenomena. This, indeed, suits well the justified interest in scientists as thinking people. Pedagogy, for instance, is essentially dealing with how to teach how to think as scientists. Scientists think, observe, test, etc. They produce their own mental representations. Then, they communicate their results: they produce scientific papers, talks, lectures, etc. They produce public representations. These public representations may then spread or not in the cultural environment of scientists, and they may constitute input to further scientific thinking. Such a chain of events is a social cognitive causal chain, and it probably constitutes a part of some a cultural cognitive causal chain. The epidemiological frame for understanding culture renders well the constant presence of the thinking scientist, whose cognitive processes are essential links in the chains described above. By contrast, most social or cultural studies of science tend to give little place to the thinking scientist, and theorise as if the cultural and social environment could produce knowledge without the intervention of human thoughts. Instead of being satisfied with macro-social explanations, the epidemiological framework requires peering into the micro-processes out of which cultural phenomena emerge. It requires peering into scientists' minds.

6.2.2 MECHANISMS DISTRIBUTING SCIENTIFIC REPRESENTATIONS

Elements of answer have been put forward as to the causes of stabilisation of scientific beliefs. Sperber (1996a, p. 92–97) compares the mechanisms of distribution of scientific beliefs with the mechanisms of distribution of intuitive beliefs, political beliefs and myths. Intuitive beliefs are acquired in the course of ordinary interaction with the environment and with others; they need no conscious learning; they mostly rely on innate dispositions. Intuitive beliefs are produced by the perception of things and events in the environment, which are processed as things and events and not as symbols or public representations. Nonetheless, intuitive beliefs can also be acquired by proxy, through communication, because communicated representations can trigger the same modules as perceptions. Non-intuitive beliefs are reflective beliefs: they are beliefs mediated by some representations; they are representations held within some meta-representation that furnishes a validating context. Reflective beliefs are consciously held. The distribution of such beliefs essentially relies on communication and the representations are often deliberately spread. This is the case of religious, political and scientific beliefs. However, these three kinds of beliefs still have different mechanisms of distribution. The distribution of political beliefs importantly depends on environmental factors: the beliefs must be relevant for a given social situation and have practical implications in that situation. The distribution of a myth importantly depends on psychological factors: myths must be attractive and memorable. What about scientific beliefs? Sperber remarks that scientific beliefs are usually hard to understand and often require important background knowledge. But once understood, scientific beliefs are normally accepted as true. Consequently, the mechanisms of distribution of scientific beliefs heavily rely on environmental, institutional, factors that provide the required education and the incentives for understanding scientific representations. But these mechanisms also importantly rely on psychological factors, since understanding is a psychological process that conditions whether scientists will believe or not the representations. Of course, there are important complications that should enter Sperber's simplified account. For instance, the act of believing scientific representations to be true often enters the process of understanding scientific beliefs. This happens especially in learning contexts, where understanding a scientific theory means understanding why it is true. Also, it is not sure that scientific thinking always requires a full understanding of the representations manipulated. Scientists can for instance wholly defer to their colleagues concerning the meaning and truth of some of the representations they use. However, the cultural cognitive causal chains that stabilise scientific beliefs do include the elements mentioned by Sperber: understanding and institutional support. The support of scientific institutions (from Universities to scientific presses) for the distribution of a representation is given only in cases where the belief is deemed to be true; scientific representations are deemed to be true only if some expert scientists who have understood the representation — or are thought to have sufficiently understood the representation to emit a reliable semantic evaluation — have accepted these representations as true. So understanding seems to be implicated at some key points in the cultural cognitive causal chain of science making, and appears as a precondition for distribution. This, indeed, contrasts with religious beliefs where understanding seems to set lesser constraints. In some cases, the content of the religious beliefs is explicitly said to be accessible to some non-human entities only. More generally, Sperber (1975) has argued that religious beliefs draw their appeal from the fact that they are forever interpretable and can be understood in many different ways, with the consequence that they can be made relevant in different contexts. Scientists tend to proscribe multiple interpretations of scientific representations.

The above described mechanisms of distribution differ because the representations being distributed differ. It is not just that the representations have different contents; it is also that they are processed differently by the human cognitive apparatus. Why is that so? What are the differentiating properties of these representations that trigger different mental mechanisms? Sperber's account reveals two important characteristic of scientific representations: they are reflective rather than intuitive representations, and they have a power of conviction that acts when the representations are understood.

Scientific representations do not constitute the taken-for-granted often unconscious understanding of the world that inform all our actions, such as avoiding bumping into things (informed by our intuitive beliefs about material objects) and managing social interactions (informed by our intuitive beliefs about people as having thoughts and desires). Scientific representations are not intuitive representations, and yet they are representations about the world (with the possible exception of mathematical representations). Both inductivist and Bacherlardian philosophies of science insist on the important distinction between science and common sense, where common sense is best interpreted as being equivalent to intuitive beliefs and intuitive thinking. Atran (1990) criticises these philosophies for assuming that scientific progress is possible only by getting rid of common sense. The inductivist tradition, as the intellectualist tradition in anthropology, argues that common sense and pre-scientific thinking (mythical and religious thinking) are limited, imperfect and tentative attempts to understand the world. These thoughts have the intention to be rational, but they fail to be so because they assert more than they can, given the limited data to which they have access. Science gets rid of the approximation of common sense. The Bachelardian tradition, as the psycho-social tradition of Durkheim and Levy-Bruhl, asserts that scientific thinking is radically different from common sense and "primitive" thinking. Scientific thinking begins with the negation of common sense and mystical beliefs. Against these views, Atran emphasizes the role of common sense as a basis of scientific thinking:

> To get from the familiar to a factual understanding of the unfamiliar does not appear to require so much a radical break with common sense as a sustained development of privileged cognitive tendencies that permit the elaboration of certain scientific ideas. In contrast to basic common-sense dispositions, these tendencies are apparently not so rigidly structured, nor so spontaneously formed, nor perhaps even so tied to specific cognitive domains. But humankind may nevertheless be universally susceptible to comprehend and elaborate them to various degrees, because of their relatively favored relationships with basic dispositions. (Atran, 1990, p. 150)

In other words, scientific thinking is distinct from common sense, but it importantly relies on it. Atran's book — *Cognitive Foundations of Natural History: Towards an Anthropology of Science* (1990) — wants to demonstrate, with the case of natural history, how our intuitive beliefs inform scientific thinking:

> my aim is to show how our universally held conception of the living world is both historically prior to, and psychologically necessary for, any scientific — or symbolic — elaboration of that world. (p. 13)

Atran first singles out a universal cognitive ability, a mental module, which produces intuitive beliefs about living entities. This modular ability is called folk-biology. Then, Atran studies the role of this ability in the development of scientific biology and natural history. His conclusions are:

> The scenario that I have explored and defended so far comes to this: Some fields of science have a phenomenal basis. The basis of such a field may be universal and depend upon a specific cognitive domain. Concern with evaluating this basis constitutes much of the initial phases in the field's development. At later stages, problems of knowing in the field take precedence over issues pertaining to its cognitive appreciation. Epistemology divorces psychology: understanding how the phenomenal basis is constructed becomes a manner of regulating inquiry in the field, but no longer constitutes its objects. At least this appears to be the case for natural history. (p. 252)

Thus, Atran explains the important relations that hold between intuitive and scientific representations. Scientific representations are semantic evaluations of some intuitive representations (the representations output of modular folk-biology in the case of natural history) specifying the limits of their content. Scientific representations are also regulated interpretations of the intuitive representations that form the phenomenal basis of inquiries (fieldwork or experiments). Science, says Atran, "goes beyond ordinary knowledge by a selective and nonarbitrary development of basic common-sense dispositions" (p. 318, note 32 in chap. 9). This development is done through "second-order representations of the first-order concepts generated by basic dispositions" (p. 249). Second-order representations are representations of representations, or meta-representations. So the pervasive presence of intuitive representations, or common sense, in scientific thinking stems from the fact that scientific representations are meta-representations with embedded intuitive representations. Atran gives the example of the contemporary natural historians for whom 'tree' does not correspond to any scientific category. 'Tree,' however, is an intuitive representation that is an output of the folk-biology module; the scientists' knowledge that 'tree' does not carve the world at its joints, does not inhibit her folk-biology module from producing and processing intuitive representations of trees. What does the scientists do with these cognitive processes and mental representations? Is he simply ignoring his own intuitive, non-scientific, representations? No, argues Atran. In the field, the natural historian still makes use of his perception of trees to make sense his environment. It is only after further thinking that the natural historian thinks of his own representations of trees so as to re-interpret them — meta-represent them — within the framework of the most recent scientific theory. This picture of the scientists both using his intuitions, and at some other points, questioning them, is also held by Barnes et al. (1996), who say:

> The machinery involved in the perception and recognition of things hums along undisturbed much of the time. For the individuals in a given culture it usually hums along in unison; indeed it has so to do for the culture to exist. The fact of its existence depends upon a certain blind conformity in perception, understanding and judgment, in *initial* responses to things. But the machinery of perception and recognition is nonetheless subordinate to reflection and calculation. The basis of sociability, and thus of humanity, lies in our shared tendencies to automatism, but its actual achievement lies in the calculative exploitation of these tendencies. (p. 127)

To this, I specify that "initial responses" that "hums along in unison" are grounded, at some point (i.e. after possibly learned connexion), in the universal nature of the mind, and, more precisely, in the human evolved cognitive abilities. I also specify that "reflexion and calculation" are implemented by specific types of mental representations, viz. meta-representations.

6.3 Some psychological principles of scientific evolution

6.3.1 Evaluating and routing intuitive representations

The fact that scientific beliefs are reflective beliefs, i.e. believed through metarepresented semantic evaluations, makes scientific thinking akin to symbolic thinking in general. Symbolism is a theoretical notion that has been developed in anthropology in order to account for the existence of beliefs that seem to have no practical value and no empirical grounds. Theories of symbolism assert that these beliefs have metaphorical meaning, which need not be explicit. Sperber (1975) has specified the form that symbolic beliefs take: they involve meta-representations, since things and events are meta-represented as having meaning, and symbolic beliefs, which are held as true, involve a validating context. As symbolic thinking, scientific thinking is based on reflective beliefs, it produces reflective beliefs, and it makes extensive use of metaphors and analogical thinking (Brown, 2003; Nersessian, 1999). But contrary to scientific beliefs, symbolic representations owe their success to the possibility of understanding the meaning of the symbols in many ways, with the consequence that symbols can be made relevant in many different situations. The hypothesis, therefore, is that scientific and symbolic thinking differ in their use of metaphors and analogies. Atran's characterisation of science is geared on the fact that scientists constrain, rather than open, how scientific representations can be understood. He says (1990, p. 12):

> Contrary to mystical analogy, the goal of scientific analogy is ultimately to reduce itself to "dead metaphor," not to produce eternally open-ended "truth." It aims to ultimately terminate any metaphorical imprecision by (ideally) accommodating one subject to another in determinate manner.

Mystical analogy is used in symbolic thinking, which is contrasted to scientific thinking:

> The goal of symbolism, unlike that of science, is not to extend factual knowledge, resolve phenomenal paradoxes or increasingly restrict the scope of interesting conceptual puzzles. Instead, symbolism goes the way of eternal truth, and is sustained in that path by faith in the authority of those charged with the task of continually reinterpreting the truth and fitting

> it into new circumstances. To the contrary, science assiduously searches out falsity in order to eliminate unknowns.

Formative analogy, characterised as *determinate* links between source and target domains, and acceptation based on understanding are two features that distinguish, according to Atran and Sperber, scientific beliefs from other reflective beliefs. These two features cannot constitute a psychological demarcation criterion between science and non-science, for the links created in analogies and metaphors are always more or less determinate, and scientific representations are always more or less understood. The fact that metaphors always offer more or less determinate links can be seen in the history of science, where the strength of analogies are tested and re-formulated as results come up — the molecular and wave theories of lights furnish well documented examples. In religious thinking also, metaphors can, according to the historical context, be more or less determinative of what can and cannot be thought. Fundamentalist religious movements, for instance, attempt to constrain as much as possible the interpretation of religious texts. Sperber thus discuss the tendencies to fix the exegesis of sacred texts, which in fact only displace the possibilities of further interpretations. The fact that scientific representations are always more or less understood is a result of their dependence on the validating context for their semantic evaluation. The validating context, indeed, can and does change with the evolution of scientific knowledge. Consequently, the meaning of scientific representations depends on other beliefs that may evolve. Lakatos (1976) gives a historical study (in a metaphorical form) of how the meaning of 'polygon' varies as knowledge in geometry increases. But if meaning varies with the increase of knowledge, then it is never possible to say that one has a full grasp of a scientific notion. In point of fact, there is nothing in the representational format that distinguishes scientific from symbolic representations. They are, at best, features that characterise better science as it has been practiced than other cultural domains, such as a lesser reliance on deference, a greater reliance on understanding and on critical thinking oriented towards semantic evaluation (see 5.4.4), and a reliance on analogies that are more 'formative' than 'symbolic'. Atran and Sperber's emphasis on the role, in science, of formative analogy and persuasion by means of understanding, is meant as an analysis of the cognitive practices that have been favoured in the history of science. Cognitive psychology as well as social-history can help characterising these practices. The characterisation of these historical practices, however, has no normative and evaluative component. In particular, it does not act as demarcation criterion. In fact, I doubt psychology could provide one : since there is no specific capacity for scientific thinking, psychology cannot tell apart scientific thinking from non-scientific thinking, thoughts that are issued by 'cognitive processes for science' from thoughts that are issued from other cognitive processes. Scientists and the community at large, rather than psychologists or psychologically informed philosophers, decide, along the history of science, what science is, and which practices and thoughts are scientific.

I submit that formative analogy and persuasion by means of understanding are possible by a regulated use of mental modules' inferential power. Analogies in science have the effect of selecting mental modules, which are put to work for the processing of a specified range of representations. More generally, scientific thinking implicates attributing new cognitive functions to cognitive abilities. This is what Atran's history of natural history suggests with the case of the regulated exploitation of folk-biology inferential power. Likewise, analogies developed by, say, the corpuscular theory of light enables using the Naïve-mechanics module for making inferences about light. Such reflective use of mental abilities through metarepresentations representing semantic evaluation and analogies is what makes scientific thinking go beyond our own cognitive limits. Yet, this thinking is done with the same cognitive abilities that have common sense, or naïve knowledge, as output. In a comparable way, Lakoff has argued that "conceptual metaphors" are part of our system of thought. They have a particularly central role in abstract thought, which is said to be a consequence of systematic layering of metaphors upon metaphors. The primary function of metaphors, indeed, "is to allow us to reason about relatively abstract domains using the inferential structure of relatively concrete domains" citep[p. 40]lakoff00.

Common sense and intuitive beliefs are not bypassed; they are evaluated and reflexively exploited. The result is the creation of cultural scientific domains of modules:

- Semantic evaluations can specify which input issue true knowledge when processed by a module. These evaluations consequently remove from the scientific cultural domain of the module many inputs that nonetheless belong to its actual domain. This is the case of trees, which issue classificatory beliefs that have no scientific significance, and this is the case of the many intuitive beliefs that are contradicted by scientific knowledge (whales are not fish, the earth is not flat and immobile, solids include empty space, there can be a set theory where an element belongs to itself and a geometry that is not Euclidean, etc.). The semantic evaluations are meta-representations that specify in which contexts the output of a module should be trusted.

- Encyclopaedic knowledge of the form "X is Y" enable triggering the module, say M with has Y in its proper domain. The inferential power of this module is then exploited for understanding X, although X may not be initially in its proper domain. The cultural domain of M can then be said to be enriched with X; the encyclopaedic knowledge acts as a router of intuitive representations towards non initially triggered cognitive abilities.

This characterisation of scientific reasoning is to be contrasted with the axiomatic view of scientific thinking. According to this view, scientific theories are sets of axioms and scientific reasoning is the application of logical rules to the axioms. This view has been much criticised in cognitive science of science for being unable to account for scientists' actual cognitive practices (see esp. Giere, 1988). The alternative account puts "model reasoning" at the centre of scientific cognition. I suggest (without developing the idea here) that much of the literature on model-based reasoning in science could be re-interpreted as the construction of models as means to route intuitive representations and exploit the inferential power of cognitive abilities that would otherwise not be recruited. In other terms, it would be possible to situate the rich literature on model-based reasoning in science in a massive modular perspective. The above characterisation of scientific reasoning is also to be contrasted with the "dual-process" account of reasoning (Evans, 2003). The dual-process account of reasoning takes its root in the experiments of the "heuristic and bias" psychology (Kahneman et al., 1982), which observed that much reasoning was based on heuristics that could often bias the analysis and lead to 'irrational' behaviour. Yet, since rational analysis is nonetheless possible and is, in certain context, actually pursued by human agents, it is possible that some cognitive processes be dedicated to this rational analysis, while other processes are implementing heuristics. Scientific reasoning, as the archetype of rational reasoning, should therefore be essentially based on the cognitive abilities dedicated for rational analysis. The alternative account is that the working of modular abilities, which are probably implementing heuristics, is constrained and reflectively exploited through meta-representational knowledge. The consequence is that it takes meta-cognition and some knowledge (probably acquired knowledge) to do rational analysis, and that this is not done by bypassing heuristic cognitive processes.

Conclusion: Symbolic thinking, as described by Sperber (1975), leaves interpretation open. It includes symbols whose connection with intuitive beliefs is loosely specified. Scientific cognitive practices, on the contrary, include pervasive attempts to constrain interpretations. This is done by specifying the cognitive abilities that can be drawn upon and how. One consequence of this specification is that some intuitions are given a true status, which grounds further reasoning. The testability of scientific claims is at bottom based on these intuitions, which provide the phenomenal basis of scientific cognition. This basis goes through a reflective attitude that interprets the phenomenological world.

important to be aware of how metaphor choice affects our understanding of a problem"

6.3.2 Factors of attraction in scientific evolution

Psychological factors

One basic idea of the epidemiology of representations is that people's production of representations is biased *inter alia* by human cognitive abilities. This probabilistic fact has consequences for the evolution of the distribution of representations in a community: representations tend to stabilize around 'attractors', whose positions are partly determined by environmental factors and partly by psychological factors. Psychological factors are expressed within the nativist hypothesis of massive modularity (this hypothesis is not necessary for Sperber's epidemiological approach, the psychological theory is chosen for independent reasons). How are these factors of attraction expressed in scientific evolution? The hypotheses on the micro-processes involved in scientific practices — the psychological level — should help specifying how these factors operate. As the modular abilities of the mind enable scientific thinking, so do they constrain and shape it. The epidemiology of representations specifies the constraining role of the modular structure of the mind, while avoiding Kantian psychologicism. Scientific thinking is much more flexible than Kant thought because humans are endowed with meta-representational abilities that enable them to reflect on their own thoughts and forever re-assess the output of their own cognitive abilities. The structure of the human mind, however, sets factors of attraction for scientific development because scientific cognition relies on modular abilities. Scientific statements that best recruit these modular abilities will be more relevant and will stabilise more easily. By recruiting modular abilities, I mean, as explained in the first section, that the input (in our case a scientific communication) is such that it either activates modular inferential power with little preliminary efforts or that it brings new information to activated modules, thus generating numerous inferences. For instance, scientists find analogies fruitful when the source of the analogy is rich in inferential power, which is what happen when it recruits the inferential power of a module that would not normally be triggered by the target of the analogy. Analogies, therefore, will be relevant and attractive depending on how they recruit human cognitive abilities. Of course, analogies may not specify directly which are the modules recruited; they may take as source of analogy already complex representations that themselves trigger differentially modular abilities. For instance, Bohr has compared the structure of the atom to the solar system. The solar system was well understood, and this rendered the analogy informative. The basic cognitive abilities that are eventually recruited by the analogy, however, include naïve mechanics — to the extent that it applies to the solar system, thanks to our understanding of planets as solid moving objects, then it applies to atoms via Bohr's analogy. The more layers there are, the more cognitive effort needed, since the cognitive processes include passing through these layers by invoking encyclopaedic knowledge. So one can expect that such hard-to-process representations stabilise only if they issue important cognitive effect. Echoing unknowingly the principle of relevance, citetbarnes96 talk about "a principle of maximum cognitive laziness," which "other things being equal, allow what is routine to count as what is correct. But if other things are not equal, if the extra work is pragmatically justified [...] then automatic tendencies may be overridden." The added value is specified by Barnes et al as being pragmatic, goal oriented, or interested in the broadest sense of the term. This is because of their focus on scientific *action* as motivated, interested action. This makes sense in relevance theory, since goals in minds are mental representations that await to be taken as premises for inference for practical actions. If some inference is enabled by some new representation, then the representation generates cognitive effect and is thus relevant. Later on, when 'experts' are convinced of the relevance of some new hard-to-process representations, they may have the power to put institutional academic machinery to work for distribution — via publishing, teaching and so on. Arguably, the evolution of scientific knowledge implies more and more complex representations, in the sense that semantic analysis, going through the numerous layers of knowledge, occupies much of scientific reasoning before the assertion can sit on its ground of intuitive beliefs. But in the end, the phenomenal basis of scientific knowledge is, as shown by Atran, ever present in scientific thinking. Evolutionary biologists never stop thinking with the concept of tree even thought the notion is not a scientific taxon— when seeing a tree in their field work, for instance, they continue to think, but reflectively, with the 'tree' concept. Scientists do not get rid of intuitive beliefs; they only re-evaluate their significance for a true account of the phenomena investigated. As a consequence, their traces remain always present in scientific practices.

Pointing out the trace of intuitive beliefs can be done by analysing the similarities between intuitive beliefs and scientific beliefs. For instance, I describe, in the next chapter, the similarities between the calculus and our spontaneous cognitive capacities for evaluating quantities. In order to point out the similarities, one must have evidence of the existence of naive knowledge that is independant from the observed scientific cognitive practices. This methodological point is important, because science is not and should not be considered as the archetypical procuct of human 'bare' cognitive abilities. Thus, Atran (1990) spends much of his book gathering evidence across cultures for the existence of a naive biology. Moreover, the analysis of similarities of content is not sufficient: a causal account specifying how the cognitive capacities have led to the production of a cultural beliefs is needed. This can be done by locating an attractor and specifying psychological factors of attraction, then account of the evolution of scientific ideas and beliefs in term of attraction. The key role of intuitive beliefs is best represented in the taken for granted beliefs upon which scientific notions are elaborated. Given the requirement to convince others on the basis of what is communicated, complex scientific representations shall be more successful if what they take for granted is also taken for granted by the adience (i.e. the scientific community). The beliefs that are most probably taken for granted by others, scientists or not, are intuitive beliefs. Thus, scientific arguments will eventually rely on obvious unquestioned truth, and such are most intuitive beliefs. Of course, intuitive beliefs are not always, in the scientific context, taken for granted. On the contrary, questioning intuitive beliefs can be said to provide the impulse for the evolution of science. But if being intuitive is far from being a sufficient reason for being taken for granted, it is certainly close to being a necessary reason.

Scientific evolution and the displacement of cultural attractors

One must pay due recognition to environmental factors of attraction in scientific production, else one may fall into Kantian psychologicism, which denied the possibility of scientific developments that actually took place. Indeed, if innate psychological traits were the only factors of attraction, then scientific knowledge would be unlikely to evolve as it does. But attractors are displaced as cultural knowledge change. Remember that the argument is based on relevance, which determine the attractiveness of representations. Relevance is highly dependant on the available information that has been acquired.

Relevance is specified by Sperber and Wilson (1986) as being relevance of an input in a context and *to an individual.* What is relevant to one individual may not be relevant to another, depending on the many idiosyncratic cognitive attributes of an individual, including his acquired knowledge and his specific cognitive skills. Yet, the fact that humans are endowed with essentially similar cognitive abilities, and the fact that communities of people partake much beliefs and assumptions, implicates that relevance of inputs can be similar in a community. More precisely, a **cognitive environment** is determined by the set of representations that can easily be retrieved either from memory or from the external environment by a given set of people (Sperber and Wilson, 1986). People of a given culture share much of their cognitive environments; so many input representations shall appear similarly relevant to people of the same culture. The representations that have a great relevance in a given cultural context are attractive, in the sense that psychological mechanisms will favour their being remembered and used. These highly relevant representations are likely to become cultural attractors. This is all the more the case when the partaken cognitive environment is known to be partaken (mutual cognitive environment): in such cases, the communicators have the incentive to make fruitful use of the mutual cognitive environment by producing representations that will generate high cognitive effects when combined with it. It is therefore also likely that attractors change their positions as acquired scientific knowledge and other cultural beliefs change.

Because cognitive environment partly determine scientists' thoughts, they also determine scientific knowledge production. In particular, Kuhnian paradigms in the history of science constitute mutual cognitive environments. The notion of cognitive environment is more general than the notion of paradigm and can characterize non-revolutionary changes in the history of science. Also, cognitive (scientific) environments constitute paradigms only if they include representations that regulate the distribution of new representations (e.g. only papers not talking about phlogiston will be published). The fact that cognitive environments can generate cultural attractors is shown by the numerous cases where a given discovery is independently made by different scientists sharing the same cognitive environment. The cognitive environment provides interpretive frames for new scientific ideas to develop within some scientist's brain and then be accepted in the scientific community (see #cross-ref).

The relevance theory approach to scientific cognition provides some psychological credibility to some sociological analyses. For instance, sociologists of science have claimed that political beliefs and 'world views' can have determining effects on the development of scientific theories. A radical example of such claims is the causal relation hypothesised between Boyle's political interest in containing dissent and comprehending it within the Anglican Church (this was a time of great turmoil in England, from the Civil War of the 1640s to the Restoration), and his scientific theory of matter as inert (Bloor, 1982) as well as the particular method he advocated for settling scientific controversies (Shapin and Schaffer, 1985). The argument in favour of the influence of the political views on the scientific views is that the latter could be used to promote the former. That scientific beliefs can indeed promote political views, although not obvious to us, is carefully documented through sociological analysis. This justifies the identification of an interest, here a political interest, in developing the scientific theory Boyle actually developed. But why should we think that the political interest did actually play a role in scientific thinking? The classical sociological argument is to show co-variations of political interests and scientific theories, together with the observation that in each case the latter promote the former. This, however, does not show the causal relation between the political and scientific views. In order to understand this relation, one must specify what mental processes relate political beliefs to the formation of scientific beliefs, and it is exactly on this matter that criticisms have great strength: indeed this relation is denied by contemporary scientists (this was not necessarily the case in the 17th century); an epistemological rule want scientific beliefs about nature to be independent from political views, and scientists analysing their own thought processes claim to obey the epistemological rule of independency. Bloor (1982) is more explicit about the cognitive processes implied: he invokes "a coherence condition" according to which beliefs are networked through "elementary laws" (e.g. 'fire is hot'), which make them interdependant. Some of these laws are fundamental categories taken for granted in most acts of communication and justification. The homology between political and scientific beliefs is then explained because scientific beliefs promote categories that are then put to use in political thinking, and lead to the desired political conclusions. This dependency of belief formation relies on psychological assumptions that are at odd with the modular view of the mind: first, categories are better thought as domain specific knowledge stored within modules, and there is no reason to postulate the further existence of social Durkheimian categories (as shared fundamental beliefs) upon which coherence conditions would apply; then, the network of belief approach need not imply that political and scientific beliefs be connected. The consequence is that the co-variation remains unexplained. Relevance theory can provide an alternative account: if a belief which is developed in investigating natural phenomena can be shown to have some consequences on political thinking, then the belief is more relevant than a belief having no such consequences; it is therefore more appealing for because it generates more cognitive effects. The consequence is that social order generates social interests with ensuing beliefs about how to promote one's social interests (political beliefs). For any new beliefs, it is more relevant, ceteris paribus, if it has implications on political beliefs that if it has none. We have a factor of attraction towards beliefs that are socially relevant, whether the beliefs are social or not. If there is a community with similar political interests, then the factor of attraction operates on this community — a scientific belief relevant to the political interest can then constitute a cultural attractor. It remains to the sociologist to show that scientific beliefs are indeed relevant to political interests, and this is what has been done in the case of Boyle's theory of inert matter. That metaphorical thinking can operate between political beliefs and scientific beliefs is not so controversial. In my interpretation, the sociologists claim is only that the metaphor "matter needs external force to move and organize itself in the same way as the people need external authority, such as the Church, to organize" has appealed to a certain class of scientists in the 17th century England. Why is the metaphor appealing? Because it enables deriving interesting consequences about both how to control matter and how to control the people. Relevance of communications on the inert nature of matter is increased because some inferences are enabled by using easily retrieved political beliefs (lowering cognitive cost) and some further inferences can be drawn about political action (increasing cognitive effect). The especially sociological consequence of this psychological property of metaphorical thinking is that the metaphor can be appealing differentially, depending on one's social interests. Different cultural backgrounds generate different cultural attractors. This explains the co-variations and homologies observed between political interests and scientific beliefs[1].

[1] Note that I have not taken a stand on the hypothesised relation between the theory of innert matter and Boyle's political beliefs. The arguments in favour of this hypothesis lies mainly on historical data. I have taken a stand only on the psychological principles that could give a causal explanation of co-variations and homologies between political interests and scientific beliefs.

Again, that anything can provide analogical resources for a target subject is not very controversial; and that socio-cultural context determines interests, which may determine political beliefs, which can then be used as resources for analogy, is not very controversial either. But the consequence is that political interests are factors of cultural attraction that have effects on the evolution of scientific knowledge. There is however a strategy that operates, I submit, along the history of science: it consists in strongly limiting and explicitly specifying which domain and what knowledge can be relevant to scientific enquiry. In my case study (chap. 7), for instance, I show how the notion of infinity was taken on in mathematics and specified as independent from theological considerations: theology was made irrelevant. It is certainly in the interest of the scientific community to show that it is sufficiently independent so as not to be told what to do or think by other communities, political lobbying or other. The practice of constraining what can be made explicitly relevant is certainly what distinguishes what Atran calls formative and symbolic analogy. Formative analogy, as characteristic of scientific cognition, cannot explicitly draw on any salient information to increase its relevance; the inferences such analogies enable are constrained because input representations are *routed* by the analogical metarepresentations, but also because the scientific context make it so that the metareprentational 'route of cognition' is the most relevant one. In any reasoning, the context of the reasoning task and the background knowledge of the reasoning individual enter into play. In particular, when a scientist reasons logically, it is not because he abstracts the logical form of the input from its context; it is, on the contrary, because the context is such that the most relevant thing to do is to reason logically[2]. This is what Sperber et al. (1995) show in their analysis of the Wason Task: they show that results on the task can be made to conform the experimentalist expectation of good reasoning (i.e. to check the truth of a conditional statement of the form 'if A then B', check instances of A and instances of not-B) in contexts where the subject's presumption of relevance of the question lead them

[2] In a constructivist view, what really happen is that the scientific context determine a specific kind of reasoning which is then acknowledged, or not, as being logical.

to derive the proposition 'not-(A and not-B)'. Interestingly, the scientific background of the subject counts much less than the form of the task, i.e. the context of the question. I submit that much of the work of scientists is to specify the relevant context in a way that determines the cognitive processes of the audience. This assertion is nothing more than the application of relevance theory to scientific practices, as practices that aim at informative communication. In a recent talk at LSE [3], Barry Barnes was showing how the meaning of a term is being determined as a function of the interests of the scientists. The case dealt with ascription of gender in the context of the development of new human biotechnologies, which raise new questions about the means for telling apart men and women. In the discussion, Barry Smith, an analytic philosopher and pragmaticist, noted that the use of the term was simply determined by the contexts of the use of the term — in some cases it was more relevant to classify such person as a woman, and in other cases it was more relevant to classify the person as a man. Barry Smith took this observation as a counter-argument to Barnes' point about the efforts and negotiations made to determine which should be the proper application of the term. In my view, Barry Smith rather specified the cognitive mechanisms that enable scientists to constrain the meaning of scientific terms: cognitive mechanisms for communication make the most of the context for cheap and fruitful inferences. Thus scientists, as communicators, do act on the cognitive environment of scientific thinking in order to constrain reasoning with reflective representations. The negotiation, then, bears importantly on what should count as a relevant context.

In this chapter, I have presented Sperber's epidemiology of representation as a way to specify the causal relations between culture and cognition, and I have applied it to the specification of the relation between scientific cultural evolution and scientific cognition of individual scientists. I have also drawn on Sperber and Wilson's cognitive theory of communication, with the justification that doing science implies communicating with other

[3]During the first conference in Philosophy of the Social Science at LSE, 10-11 June 2005

scientists, and convincing them of the truth of one's claims. I have then attempted to show in which way the epidemiology of representations could specify some of the claims in the sociology of scientific knowledge by clarifying the underlying psychological processes; then I have speculated on these cognitive processes. In the end, the epidemiological framework can help making the parts of psychological and socio-cultural causal factors determining the evolution of science. What is the import of the Strong Programme in the epidemiological framework? One can continue to draw on the social theory it has been using for understanding knowledge production (c.f. the compatibility claims of the previous chapters) or re-assess the case studies in the light of the cognitive processes they presuppose. However, its essential import is to show the significance of the socio-cultural determination mentioned 'cut deep': acting on the evolution of scientific knowledge, they are causes of the content of our scientific knowledge.

Chapter 7

Mathematical cognition and history: a case study on the notion of infinitesimals

In a paper whose title is "The cultural and Evolutionary History of the Real Numbers" (2005) the psychologists Gallistel, Gelman and Cordes say: "Our thesis is that [the] cultural creation of the real number was a platonistic rediscovery of the underlying non-verbal system of arithmetic reasoning. The cultural history of the real number concept is the history of our learning to talk coherently about a system of reasoning with real numbers that predates our ability to talk, both phylogenetically and ontogenetically". The paper puts forward strong evidences in favour of the existence of "a common system for representing both countable and uncountable quantity by means of mental magnitudes formally equivalent to real numbers", but it actually says nothing about the cultural history, or how the 'platonistic rediscovery' happened. The quote also expresses a psychologistic philosophy of mathematics, as Mathematics as it evolved in history of the real number is claimed to be coherent talk "about a system of reasoning". Psychologism, the thesis that mathematics is about the human mind, has been strongly criticised and was officially dismissed by the arguments of Frege (1884, 1893) and Husserl (1900). Their main point was that the truths of logic are objective and independent of psychological empirical and subjective facts. Psychology deals with what people believe to be true while logic deals with what is necessarily true.

The study of Mathematical abilities generates a problem: how to study the cognitive bases of mathematics without being psychologistic? It also generates a research question: what is the role of these mathematical abilities in the historical evolution of mathematics? The hypothesis that mathematical truths reflect the universal structure of our cerebral representation still appeals to many psychologists and philosophers of mathematics. The good idea behind Gallistel et al.'s quote is that the history of ideas, including the history of mathematics, is importantly determined by aspects of the human mind. This is the idea I want to pursue in this chapter. But a prerequisite of this research programme is that much more must be said on the social processes through which aspects of the human mind help determine cultural evolution. Showing a similarity between mental abilities and mathematical theories, together with the anteriority of mental abilities, strongly suggests a causal relation between cultural ideas and these mental abilities. But naturalistic studies must also specify through which causal processes the similarity arises. Are Gallistel et al. psychologising mathematical theories, when they posit the existence of a "system of reasoning with real numbers" in our head?

In this chapter, I attempt to specify the relation between mathematical abilities, as those recently hypothesised by cognitive psychologists, and the history of mathematics, seen as a cultural product. I thus reiterate the claims of the previous chapters with the special case of mathematics and present an epidemidemiological analysis of a mathematical representations — the notion of infinitesimal.

In the first section of this chapter, I analyse why many enquiries into the cognitive foundations of mathematics turn out to be psychologistic, i.e. assert that the truths of mathematics and logic are psychological facts. I then attempt to show what philosophical presumptions on the nature of mathematics lead to psychologism. I argue that the only way out of the problem is to acknowledge the importance of the social aspects of mathematical practice, especially the implementation of norms. I then present the epidemiology of representation as a way to develop non-psychologistic enquiries into the cognitive foundations of mathematics and its evolution. In the second section of the paper, I give a brief account of the psychological studies on the human ability to perform arithmetic reasoning — the number sense. I point out how the number sense can have causal effects on the respective distributions of Leibnizian and Newtonian ideas about the infinitesimal calculus. This causal effect is explained in terms of difference of relevance of the two concurrent ideas to the mathematicians of the 18th and 19th century. In the third section of this chapter, I attempt to track down mathematical representations of the infinitesimal calculus, as they occurred at the turn of the 17th century France. This analysis aims at pointing out the richness of the events that constitute the evolution of mathematical knowledge. In particular, I advance historical evidence in favour of the existence of a cultural attractor towards mathematical notions that resemble the notion of limit.

7.1 Psychologism and the cognitive foundations of mathematics

7.1.1 Paths to psychologism

Psychologism, in its crude form, is the doctrine that asserts that "logic is a study of the mind" (Macnamara, 1986, p. 10), and more generally, that mathematical principles are principles of the mind. Since Frege and Husserl the question seemed to be settled: psychologism is an erroneous philosophical theory of mathematics. Nonetheless, I will argue that psychologism is still a lively philosophical problem. In particular, the use of logic when accounting for rational behaviour, especially the rational behaviour of mathematicians, has led some authors to develop a new kind of psychologism. Cognitive science asserts that human behaviour stems from cognitive processes. Applied to Mathematics, this means that the production of proofs and mathematical concepts should be explained in terms of cognitive processes. Moreover, one important paradigm in cognitive science asserts that cognitive events are performances rendered possible thanks to some cognitive competences. These competences can be domain specific and perform specific tasks, through specified computations on mental representations. Within this framework, it is natural and fruitful to hypothesise the existence of cognitive competencies such as a 'logic module' and/or an 'arithmetical module', which are mental devices that produce/perform logical and/or arithmetical reasoning. Yet, once this assumption is made, the threat of psychologism is not far: cannot Mathematics be reduced to the proper functioning of mental mathematical abilities? And if not, what else determine the content of mathematical knowledge?

Macnamara's theory of logical knowledge provides a good example of a thoughtful psychological study of mathematical and reasoning abilities, which lead to some kind of psychologism in spite of the author's denial. In "A Border dispute" Macnamara (1986) called for a research program based on the idea that the mind contains some innate devices from which originate our reasoning and basic logical skills. The goal of the program was to discover how these devices work. Logic, and more precisely the logic of type, was chosen as the appropriate, and even essential, mathematical tool for the study of these reasoning and logical abilities (Macnamara and Reyes, 1994). According to Macnamara, our basic logical skills rely on a "mental logic", which accounts for our linguistic resources of expression and understanding, and our ability to grasp inferences. More generally, "The mind in part of its functioning applies the principles of that [mental] logic". The mental logic is in correspondence with "each ideal logic (true to intuition)" (1986, p. 22) and includes fundamental principles such as the principle of contradiction. Logical competence is error free and "gives rise to intuition of absolute necessity" (1986, p. 28). It constitutes a 'competence', as opposed to 'performance', following Chomsky's distinction in his theory of universal grammar. This means that the mental logic, or logical competence, is not framing all our thoughts as in a Kantian theory (sometimes called transcendental psychologism). The mental logic constitutes an aptitude that we can, and must, call on in order to perform good reasoning. The main purpose of this distinction is to allow the possibility of logical errors in the performance of logical tasks. The logical competence "abstracts from logical error, from other psychological functioning that accompanies logical thought, and from the specifics of the many devices that could apply the competence" (1986, p. 27). In that way, Macnamara wants to account for the facts that laymen have the same ideals (in their behaviour) as those implemented in logic (e.g. consistency), and that formal logic is based on basic intuitions.

Macnamara defends his theory against the accusation of being psychologistic. His claims, he says, "have to do with access to logical principles, not with justifying them" (1986:42), the latter being the work of logicians. What does Macnamara mean by "access to logical principles"? Is mental logic a kind of ladder which gives access to the objective realm of logic? In that case the truth is already there; mathematicians describe it and psychologists describe how and why the description is possible. Macnamara, however, explains the human possibility of doing logic by hypothesising that the basic principles of logic are hardwired in the human mind. When he considers our access to logical connectives he merely asserts that logical connectives are already in our minds. The use we make of connectives is the result of the activation of our logical competence, the mental logic. Hence Macnamara describes the access to logical principles as nothing but the working of our mental logical competence. There is therefore no ladder to the realm of logic; we do not observe this realm that is supposed to be independent from our psychological abilities. The ensuing image is that of a human made logic that mirrors an objective realm of logic, but is independently constructed. The reason why there is an identity between the two logics is not explained, but the realm of logic is nonetheless used as guarantying and objectifying the human made logic: it provides the semantic referent of logics, even though there is no causal relation between the objective realm of logic and the human made logic. In these conditions, the 'realm of logic' is a post hoc entity with no role in the explanation of the evolution of mathematical knowledge. Still, Macnamara continues his defence by softening the meaning of 'access to logical principles'. Mental logic, he says, does not *generate* logic; it only *assesses* its validity. But if the basic logical competence is just the cause of our conviction, then the assertion is just that we have a feeling of certainty because we have an innately determined feeling of certainty. This does not provide, as Macnamara claims, "the key element in the psychology of human reasoning". For this, Macnamara's mental logic must be generative; it must, in particular, enable the production of formal logic. One possibility to combine the assessment procedure and the generative requirement is to postulate the existence of a device that generates some logical-like propositions, upon which the logical mental module would operate a selection. But here again logic finds its justification in psychological facts, namely passing the assessment tests of mental logic. Logic is in Macnamara's theory, the very result of our (ideal) performing of the logical competence. The truths and the laws of logic can be reduced to laws of psychology because the formers are just the expression of some characteristics of our mind. But such characterisations of our mind are actually laws of psychology. In brief, to give a mental reality to the laws of logic implies that the objectivity and the normative character of logic stem from the laws of thought. Macnamara's appeal to Platonism does not provide a way out of psychologism, since no epistemic relation between the platonistic realm and our logical knowledge is specified. But Macnamara's epistemology does not include an account of how we refer to, or have any intuition of, Platonistic entities *outside* our minds. On the contrary, his cognitive account of logical knowledge is wholly *internalist*. (The Platonistic School has, however, provided an account of how reference to Platonist entities is possible. I will come back to it shortly.)

Macnamara's theory is a moderate version of psychologism because it asserts that logic expresses properties of the mind but not in the sense of Mill that "logic is an introspective science generalising over inferences that are judged necessary" (Macnamara 1986, p.10). Logic is not an empirical science generalising over people's reasoning. It is psychology that must contain logic as an *a priori* science. Thus Macnamara comes back to a theory that looks very much like Kant's 'transcendental psychology.' Meanwhile, his competence/performance distinction enables him to avoid the difficulties of a strict Kantien theory with regard to the possibility of logical error. Kantian transcendantalism plus the competence/performance distinction enable Macnamara to avoid much of the strong criticisms of

Frege and Husserl against psychologism. Yet, if the relation between logical theories and logical abilities is not further specified, the temptation of psychologism is not far, as logical theories are said to be nothing more than the expression of properly functioning mental abilities.

Other mathematical competencies have been posited by psychologists. The cognitive foundations of arithmetic have in particular been the object of interesting psychological experiments and theories. Based on animals and pre-linguistic children ability to distinguish different quantities, psychologists have asserted the existence of a 'real number system in the brain' (Dehaene, 1999; Gallistel and Gelman, 2000). As illustrated by the quote in the introduction, from Gallistel et al., psychologistic philosophy of mathematics threaten also this domain of mathematical knowledge. I now pass over this topic, because the two following sections will be dedicated to the relations between arithmetical abilities and mathematical knowledge of arithmetic.

Of course, not every psychologist is convinced of the existence of mental abilities dedicated to mathematical reasoning (e.g. Johnson-Laird and Byrne, 1991). My present concern, however, bears not so much on the specific capacities psychologists posit, as on the relation between these capacities and Mathematics. In fact, I think that these psychological investigations and theories about mathematical abilities are mostly right. Yet, for those theories *not* to be disconfirmed by the arguments already raised by Frege and Husserl against psychologism, the role of these capacities in the making of mathematical knowledge must be clarified. For instance, Dehaene's assertions seem wise and modest: he urges teachers of mathematics not to discard basic intuitions in their teaching, with the argument that such mathematical intuitions do exist and should play a role when learning mathematics. But other scientists have made stronger and rather bold assertions. For instance, the logician Krivine, supported by the philosopher J. Petitot, asserts in a French journal for the popularisation of science (Science et Vie (2002)—titled 'Intelligence reveals its true nature'), that mathematical theorems are nothing but discoveries of the functioning of our mind. Godel's first incompleteness theorem, for example, would be the mathematical discovery of our sleeping program [1].

Penelope Maddy asserts that we “possess intuitive, non-linguistic knowledge of general facts about sets, and intuitive principles like the simpler axioms and the iterative conception are justified by their accuracy in formulating this intuitive knowledge” (1980, p. 189). In order to save Godel’s Platonism from the epistemologist’s criticism (how do we access the Platonic world of mathematical entities?), she describes mathematical intuition in terms of neurophysiological processes (Maddy, 1990). Maddy, here, is very close to psychologistic theories. Yet, with Godel, she asserts that Mathematical concepts do refer to things outside our minds, which are (contrary to Godel and in accordance with Quine and Putnam) in the physical world. This, if it was true, should save her from psychologism. Unfortunately, as she acknowledges herself later (1996) (The later ‘naturalist Maddy’ criticises the previous ‘realist Maddy’; in this chapter, I normally refer to the realist Maddy), this does not square with the practice of mathematics: mathematicians do not intend to refer to things in the physical world and, more importantly, they do not call on how the world is to justify their claims. It is, however, another assertion that actually makes Maddy’s account not psychologistic. She says: “[mathematical] intuitions can be false, so no matter how obvious they seem, they must be confirmed like any theory, and like any theory, they can be overthrown”. Intuitions do play a major role in mathematics, but they do not have the last word. This protects Maddy’s theory from the criticism against psychologism but let us with another problem: if mathematical intuitions are not sufficient for assessing the truth of mathematics, what is?

7.1.2 Where is then the norm? Strategies for avoiding psychologism

The kernel of the refutation of psychologism relies on the essentially normative component of mathematical practice. In mathematics, we just cannot say whatever we want, even if it corresponds to our personal strongest

[1] the hypothesis, although widely speculative, is based on the relations that exists between programs and proofs; in particular, there is a relation between the proof of Godel’s incompleteness theorem and recovery algorithms

intuition. The problem with psychologism, is that it cannot account for this fact. In the setting of competence theory, the refutation of psychologism implies that psychologists cannot designate a class of mental processes that would necessarily produce mathematical truths. At most, cognitive processes sustain and allow definitions to have some content and proof procedures to be applied. Mathematics is normative and norms differ from competence:

- A standard of justification is something external. It is not subjective.
- Justifications are performance. It is these performances that are the object of assessment (truth, coherence, ...), not the underlying competence.
- We cannot give an instance of the right cognitive ability which is to be used when doing mathematics. Doing mathematics does not consist in categorising neurological events.

On the other hand, we do want to give a role to mathematical intuition. At some point, formal justification must end and give place to mere feelings of certainty. So mathematicians *do use* their intuitions to do Mathematics, even if mathematics is *not about* these intuitions. Taking into account the role of basic mathematical intuitions in explaining mathematical practices is recognising that mathematicians thinks with cognitive devices (their brains) that are already loaded with content and pre-specified processes (as representational bases of intuitions and inconscious inferences).

There are various strategies for recognising the normative aspects of mathematics, and thus avoiding psychologism, and at the same time giving a proper role to cognitive processes. Contemporary Platonism, Modesty and the cognitively informed Sociologism that I will advocate, are such strategies [2].

[2]Formalism, however, is not such a strategy, since it denies as much as possible the role of intuitions. Intuitionism and logicism, or set-theoretic fundationalism, acknowledge the role of only a restricted portion of the intuitions at work in mathematical practice. In any case, it is the status of these intuitions as the cognitive grounds of mathematical thinking that is at stake, rather than the question of which are the most fundamental

Much of my criticism of Platonism merely consists in saying that Platonism does not permit to meet the epistemological challenge (i.e. how do we access Platonistic entities?) Godel's Platonism, for instance, does not provide a clear account of the intuitive contact between us and mathematical objects. More recent attempts to give such an account have developed what Maddy (1989) calls "compromise Platonism" and assert that mathematical entities are in the world. Two alleged links between the worldly mathematical realm and mathematicians are put forward. First, Maddy (1980) provides an account on how mathematicians refer to sets: along the line of the causal theory of reference, she shows that we perceive sets and that this perception is at the origin of our intuitive knowledge about sets. Maddy forcefully describes a referential concept of set that is developed through interactions with sets of real things. This referential concept, she says, remains a *naïve* concept. It is certainly the one we use in our day-to-day reasoning about sets of things and that may be at the basis of our naïve understanding of numbers. But to which extent, and how, do set-theorists use the very same concept in their reasoning? In any case, a mathematician is not allowed to prove a theorem by referring to instantiated sets of real things. This casts serious doubt on the referential character of mathematical concepts. The naïve concept of set does play a role in Mathematics, but it is not the one of allowing mathematicians to refer to worldly things. Mathematical notions have a referential character before (naïve theories) and after (applied mathematics) mathematics proper, but mathematical practice allows no such reference. The second alleged link between the world and mathematical knowledge calls on evolutionary theories. It is mentioned by both Maddy and Macnamara, Dehaene (2005) appeals to it, and Mercier (2006) develops the idea. Innate mathematical concepts are the fruit of an evolutionary adaptation of our cognitive apparatus to the world. These concepts could not be misleading, or else they would not survive evolutionary selection. Therefore they are concepts providing true

intuitions, or which are the intuitions most proper for developing mathematics. So I remain with the task of showing that my strategy, cognitively informed sociologism, is better than Platonism and Modesty.

intuitive beliefs. The problem is that evolutionary selection does not guaranty truth. As Maddy acknowledges herself, intuitive beliefs may be false. Suppose, however, that mathematical intuition happened to be true intuitive beliefs, it would remain that in lack of any supplementary means for assessing the truth of a mathematical proposition, we are bound to rely on psychological facts only. The causal link does not cancel psychologism, it only partially suggests causes of the "unreasonable effectiveness of mathematics in the natural sciences" (title of Wigner's article, 1960). Mathematical knowledge is not directly knowledge about the world, since the truth of its propositions is not assessed on the ground of what the world is. Macnamara's and Maddy's intuitions are mathematical intuitions, but they are not intuitions of the Platonistic mathematical realm, whether instantiated in the world or not. Platonism does not explain the processes through which we assert the truth and the falsity of mathematical propositions.

Any purely psychological account of mathematical production leads to psychologism, and psychologism downplays the normative nature of reasoning. This led some philosophers such as Davidson to urge for *modesty*: norms cannot be explained away in a naturalistic framework. As a consequence, Mathematical practice cannot be reduced to its causal antecedents. In this line of thoughts, Engel (1989) endeavours to provide a descriptive theory, which would grasp human rationality. The research program consists in finding the laws that a person would ideally follow; it assumes that people are rational and it projects to abstract this rationality. Engel compares this abstraction with the abstractions made in physics (e.g. a system is assumed closed) or in economics (e.g. with the homo-economicus). The abstraction here is made on the empirical limitation weighting on reasoning, and the idealisation is therefore not an empirical generalisation. What is taken into consideration is the intuitions we have of the validity of arguments in order to build a logic that suits them as well as possible. The research project is therefore to describe good common sense. There could be two strategies for building this logic (1989, p. 397). The first one is to gather our particular intuitions of validity and from this build a system of logic that accounts for those intuitions. The second one, which seems more feasible, is to start from an already existing logic (e.g. first order classical logic) and adjust it to our intuitions by making some corrections, possibly loosening the requirements for validity. An obvious adjustment, for instance, when starting with first order logic is the one required for the connective 'if ... then' which is notorious for giving some counterintuitive results (1989, p. 44). The idealisation can be either maximal, which implies that we suppose a strong rationality and build a strongly normative logic, or minimal, which means that our requirements in the consistency are loosened in order to account for more rational actions. The first choice answers to the demand that logic be normative and guide precisely our reasoning, the second choice answers to the demand that logic be applicable and correspond to people's actual rationality. The maximal idealisation would account only for what a completely and perfectly rational person would do, but it is not realist to ascribe such rationality to people. Conversely, one wants to avoid the danger of just giving a description of how people generally act, and thus failing to account for rationality. Engel advises to adopt the principle of thoughtful equilibrium between the empirical and normative constraints. This shall take into account both rationality and the restrictions set by cognitive aptitudes. The logic accordingly constructed shall give rules as close as possible to our actual reasoning. It would give an "empirical theory of deductive competence" (1989, p. 392). The logical system thus discovered is justified empirically by judgements of acceptability on particular inferences, but its propositions are justified by the system, as in any formal logic. We still deal with a deductive logic and we cannot be accused of justifying inferences with psychological facts.

The first striking point is the similarity between Macnamara's and Engel's research programs. They both want to account for natural reasoning on the ground of a logical construction, and their research programs consist in discovering the appropriate logic, which shall be in accord to our logical intuitions. The logical incompetence of actors is attributed to performance factors such as the memory space and the ambiguous interpretations of data. So the difference between the two projects may be the status which is given to the logic accordingly constructed. Macnamara asserts the existence of this logic in our mind in the form of mental devices. Engel is much more careful on this point. He must face the following dilemma: either he gives a causal status to the empirical theory of deductive competence and then he commits the 'sin' of psychologism; or he does not give this causal status and then loose the explanatory power of the theory. Engel has a balanced discourse with regard to the psychological reality of a system of logic, once characterising it as respectable, though improbable, empirical hypothesis (p. 413), and elsewhere stigmatising it as being an "illusion descriptive" (p. 393). His stance tends to be neutral with regard to the constitution of the mind. Yet, Engel assumes that, notwithstanding this neutrality, his research program is relevant to psychology. Why is that so? If the relevance is not in establishing some causal laws describing our mental processes in reasoning, where does it stem from? How does this logic contribute to the psychology of reasoning? As a matter of fact, Engel cannot attain the "thoughtful equilibrium" he is aiming at. While the *content* of his logic can be a compromise between 'strongly normative' and 'descriptive of people's actual behaviour', the *status* of the resulting logic cannot. It is either the logic of the psychologist, and then one should assume that it is a description of some thought process, or the logic of the logician, and then it is purely normative. The dilemma strikes once more: it seems impossible to describe mathematical cognitive abilities with mathematics, reasoning skills with logic, without falling into psychologism.

The strategy I advocate is a way out of this dilemma: humans obviously have the ability to do mathematics, yet this ability cannot explain the normative component of mathematical knowledge. As naturalists, the obvious move for continuing our investigations is to come back to the empirical data. What is this mysterious normative component that forbids us to describe cognitive processes as the cause of mathematical knowledge? Most of us have experienced a feeling of certainty being contradicted by some epistemic authority, and the mistake of mathematicians are discovered and qualified as such by peer mathematicians assessing the work of others. The observable data we have concerning the normative aspects of mathematics is the assessments that take place among mathematicians, and among teachers and their pupils: this data is, at first glimpse, social data.

7.1.3 Considering the social aspects of mathematical production

Here are a few of the social events that make up Mathematics: Mathematics is taught at school; this teaching creates a community of mathematicians; mathematicians submit their work to journals and editors, they communicate their results in the hope that their work be recognised as mathematics. Proof-readers and readers assess and may eventually recognise and use the work of other mathematicians. Social normative behaviour is present in all those events: a pupil is congratulated for his work, an article is accepted or not, a work is qualified as good or erroneous. Mathematics is made of assertions, and it is through a social process that assertions can enter the corpus of Mathematics. Among the infinite set of possible assertions, mathematicians choose, collectively, which are true, which are nonsensical, which are not worth considering, which can gain the status of proof, definition, axiom or theorem. Mathematical knowledge, in brief, is not the product of solitary mental devices. Psychologism is the consequence of ignoring the important role of social and historical phenomena; it is based on an incomplete understanding of mathematical production. For instance, the fact that mathematics is actually learned is a psychological fact, but the teaching itself is a sociological fact, as it involves the interaction of several individuals. Likewise, the fact that mathematicians do produce theorems is a psychological fact. But the fact that their productions are labelled (or not) theorems, thus recognized as mathematical productions by the community, is a sociological fact. The automatic application of a technique, or the blind rule following of the individual, does not furnish by itself a criterion of correct procedure, it just allows the procedure to be. Scientific and mathematical knowledge cannot be reduced to being the mere output of some mathematical abilities. The more complex story is that psychological processes determine behaviours that leads, in certain conditions, to the production of mathematical knowledge.

In graph 7.1, I present three pictures representing how mathematics is produced. The simple point I want to make is that the more complex picture is better than the simplistic picture. The picture C conveys the idea that the social factors of mathematical knowledge production should be taken into account.

Social processes, although their existence is acknowledged, are often not given their real importance. The opponents of the sociology of scientific knowledge would argue that whether recognised or not by the community, mathematics is done by individuals, it is the product of their minds. Social interactions, they would go on, have no causal action on the *content* of Mathematics. At most, cultural factors may push mathematicians in one direction or another. Thus, they would say, the more complex picture says nothing more than the simplistic ones. Maybe such a view can lead to a genuine research program. It is not, however, a question that can be settled a priori: these events are real, they are constitutive of mathematical practices, and they are the observable manifestations of the normative nature of Mathematics. Mathematics is de facto a collective, historical production. A psychological reductionist program should aim at showing that the constitutive social events have no causal impact on the content of mathematics. In any case, the claim cannot be taken for granted; it is not an *a priori* truth. The psychological reductionist research programme would greatly gain if it would take the challenge set by the sociology of scientific knowledge seriously. The challenge, both empirical and theoretical, is huge. Let me point out at some of the difficulties that a reductionist program would have to face:

(1) One point of social historical studies is that even if Mathematics is indeed highly constrained by cognitive abilities, it still remains underdetermined by these abilities. Contingent, historical, social determinations come to fill the gap. The sociology of knowledge thus aims to point out the causal role of these social determinations on the content of Mathematics. A reductionist account cannot simply deny these causal roles. It should on the contrary try to show that social behaviour and outcomes of social processes are, in turn, wholly determined by our cognitive apparatus, *and* that

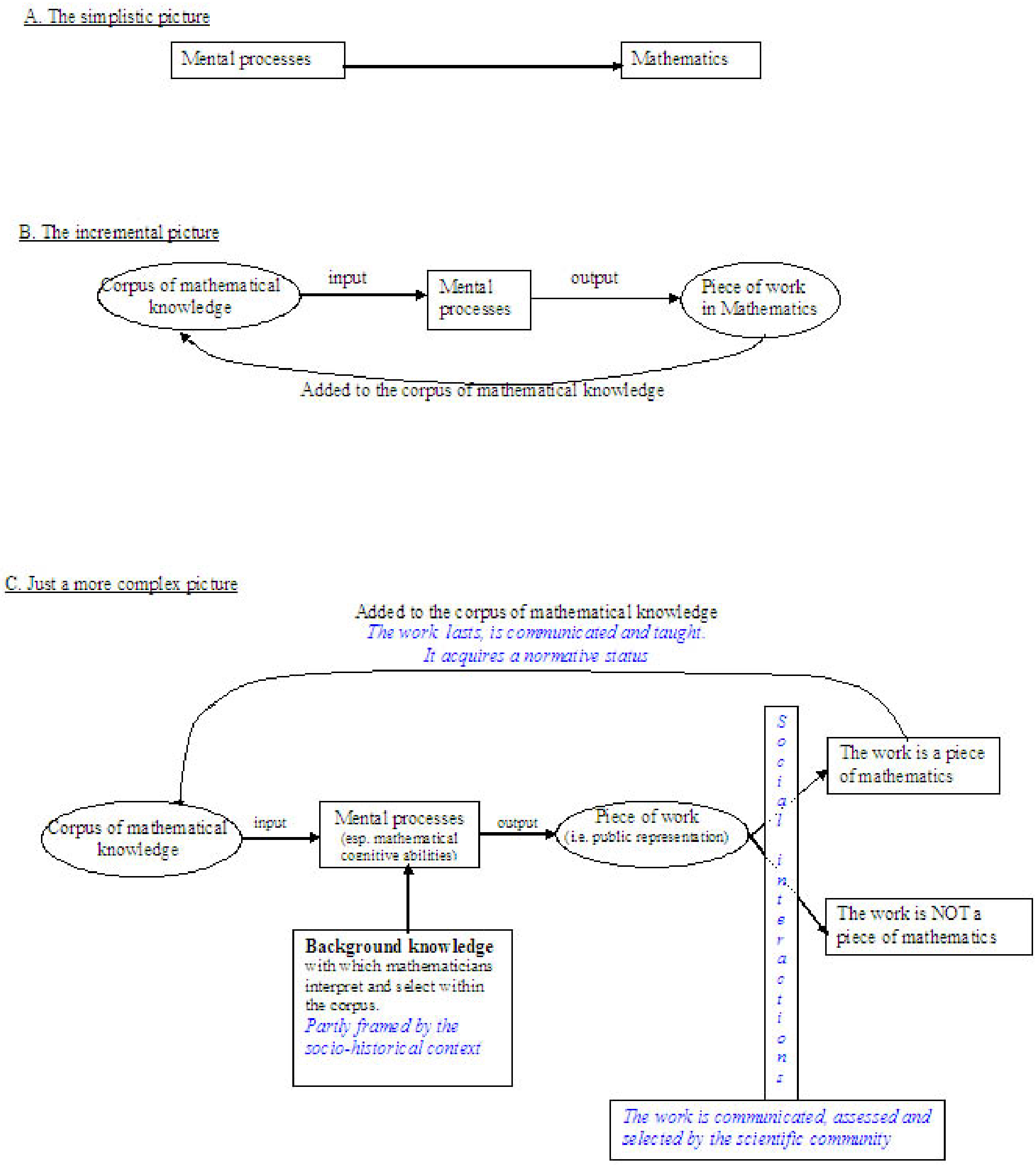

Figure 7.1: Tree views of mathematical knowledge production

this cognitive apparatus is unaffected by social contingencies. For instance the thesis that the "cultural creation of the real number was a platonistic rediscovery of the underlying non-verbal system of arithmetic reasoning" (Gallistel et al., 2005) should be confronted to historical events: Were the 18$^{\text{th}}$ century developments toward the concept of limit determined by our sole cognitive abilities (see next chapter for an analysis of this example)? What is then the significance of Non-Standard Analysis (Lakatos, 1978)? It seems that even at the most basic levels some decisions have to be taken, and that these decisions are taken collectively, by socially situated mathematicians (Bloor, 1994).

(2) The second difficulty for the reductionist program lies in the normative nature of mathematical practice. The norms are expressed in the social behaviour of assessment. Individual productions are judged through some social processes that include many more events than the ideal, unconstrained, performance of cognitive abilities. Showing that social norms are wholly determined by our cognitive apparatus seems difficult. In particular, it remains to discover the processes through which semantic properties of cognitive processes are implemented in social norms.

There is a belief that prevents the social study of mathematics: it consists in assimilating the collective production to the individual production, the former being thought as being nothing more than the sum of the latter. This mistaken assimilation is the assumption that transforms sound cognitivism into psychologism. I have argued that the best alternative theory to psychologism is a cognitively informed sociologism. Sociologists point out the historical and social contingencies at work in the making of mathematics and thus render psychological reduction difficult (e.g. Bloor, 1978). Conversely, psychologists bring up empirical evidences of the cognitive determinations that make mathematicians think the way they do, and this lessens the importance and causal power of historical and social contingencies (as Gallistel et al., 2005). There is an obvious tension between psychological and sociological theories of Mathematics. But this tension set challenges that can be fruitfully taken. It is such a challenge that I will take in the next sections: I will argue that the innate endowment of the human brain might have determined the evolution of Mathematics in one direction, while social contingent factors were pulling in another direction. More precisely, while the social situation was favouring the development of the atomistic notion of infinitesimals in the 18^{th} century France, I suggest that the concept of limit was favoured for reasons related to our innate ability to understand quantity.

7.1.4 SOME CONCLUSIONS ON THE HISTORIOGRAPHY OF MATHEMATICS

Assertions about the cognitive foundations of mathematics rarely show how these foundations have operated during the historical creation of mathematical knowledge. On the one hand, the cognitive foundations are meant to refer to psychological events that are *necessary* for the mathematical thinking that they found. Notably, if there are such things as mathematical abilities (abilities that are cognitive foundations of some mathematical thinking), then these abilities must have played a role in the history of mathematics. On the other hand, the history of mathematics cannot be viewed as the mere expression of our innately specified mathematical intuitions. The many centuries and sophisticated developments to arrive at our current mathematical theories include numerous determinative events that probably cannot be ignored in our account of knowledge production.

How to write the history of mathematics then, when having in mind the cultural, situated, practice-based aspects of mathematical thinking, and at the same time the goal to explain how human *minds* have produced mathematics? Methodological principles for the history of mathematics include 'avoid anachronism' and 'do not neglect the local causes that determine mathematicians' behaviour and, through it, the content of mathematics.' Psychologism and the so called "realist-teleological philosophy of mathematics" (Bloor, 1976) are epistemological views that tend to confine historical investigation to a history of mathematical ideas disconnected from the causal, local and situated making of mathematics. A "realist-teleological philosopher of mathematics" is happy with present-ing the history of mathematics as the necessary development of some initial notions. In this view, cultural and social analysis of mathematics is confined to enquiries of mathematicians or on how the historical 'destiny' of mathematics defeated cultural prejudices; cultural contingencies and the contingent endowment of the human mind can have no role in the development of mathematics, except the one of directing attention towards some aspects of a pre-existing mathematical realm or enabling perception of this realm. Psychologism has similar restrictions on the history of mathematics: if mathematics is nothing but the pure (unstained by social causes) expression the human mind, then social investigation and detailed history of mathematicians' choices and actions have little relevance. What are the alternatives to these handicapping and inhibiting epistemological views? Social constructivism appears as the main candidate for naturalistic history of mathematics, but it tends to ignore the determinative role of the nature of the human mind. Both sociology and psychology, I argued, should have their say in historical studies of the mathematics.

The micro-analysis of how mental representations are produced on the basis of inputs and mental processes is what can clarify the role of psychological factors in mathematical knowledge production. But the account of the historical production of mathematics cannot stop here. What is the input, and what has determined its content? What will become of the mental representations in the head of one given mathematician? He may want to communicate his ideas, but through which means? How and why can he be successful? Asking these questions is putting mental processes as events in the constitution of cultural phenomena, i.e. in complex causal chains that span minds and the environment. In particular, mental representations constitute cultural phenomena only if they are well distributed into a community. Factors of distribution involve the properties of the cognitive apparatus of the people in this community, which may include numerous local properties, such as the current interests of the people in this community and the background knowledge they use to make sense of new inputs. The physical properties of the environment are not to be neglected either: it can make a great difference if a mathematician can communicate with written symbols or just vocal utterances, if he had been using compasses or computer simulation when developing his thoughts. The geographic distribution of mathematicians can also be determinant. Finally, the institutional mechanisms of distribution of representation, teaching and research institutions, scientific journals, etc. are factors of distribution that can importantly determine what knowledge will be distributed and developed. We can understand cultural phenomena through a micro-analysis of the events that constitute causal chains out of which the cultural phenomena emerge. Sperber's epidemiology of representation provides a way to take into account both psychological and environmental factors that frame scientific and mathematical knowledge.

Sperber and Claidière (2007) list the following important "forces" of cultural evolution: they are "psychological forces," among which one can distinguish "content-based forces" and "source-based forces." The content-based forces drive cultural evolution in virtue of the content of the representations transmitted. Source-base forces determine transmission in function of the source of the information. "Psychological forces involve mental mechanism that are in part genetically determined and that are in part the output of culturally informed cognitive developments," explain Sperber and Claidière. There are also "ecological forces" that should not be neglected. I mentioned above the means of communication, the geographical repartition of the population, the available tools and natural resources. Understanding ecological forces could thus make justice to Maddy's ideas about the role of the world on framing mathematical ideas, both ontogenetically and phylogenetically, as, for instance, ecological forces include the facts that the world is such that we can manipulate sets of things. Importantly in the evolution of such complex cultural phenomenon that is the evolution of mathematical knowledge, "ecological forces involve aspects of the environment that are themselves the result of human action, and therefore of human culture" (Sperber and Claidière, 2007). From a historical perspective, psychologism commits the mistake of taking into account only one very specific force of cultural evolution: these aspects of "content-based forces" that moreover abstract away cognitive mechanisms caused by social interactions. This reduction is disconfirmed by the richness of mathematical practices. Mathematical education, for instance, is a process that involves all "forces" of cultural evolution: the buildings in which the courses are taught, the cognitive resources put at work by the students, the availability and authority of teachers, etc.

I submit that historical advents of some mathematical theory may sometimes be viewed as new ways, culturally implemented, of recruiting mathematical abilities. Mathematical abilities are put to work to solve the problems they have evolved to solve in our evolutionary environment, then, they are also put to work for some new culturally framed or created problems. For instance, the number sense is put to work for tracking preys or choosing the trees with most cherries, and it is also put to work for comparing linguistically represented quantities, for evaluating the plausibility of some calculation in a physics problem and so on. An important question about the relations between psychological mathematical abilities and the history of mathematics can then be phrased as: 'what are the cultural events that enabled recruiting mathematical abilities so as to produce mathematical knowledge?' And the answers should take the form of detailed case studies. These answers would flesh out psychologists and philosophers' assertions about the cognitive foundations of mathematical and logical thinking. How do pre-existing structures constrain mathematical production in actual cases? Can there be a naturalistic understanding of Gallistel et al.'s process of "platonistic rediscovery" (quoted in the introduction)? In particular, we want to show the causal role of psychological abilities in the making of mathematics without saying that mathematics is about these psychological abilities, or that it is their mere expression.

Advances in the psychology of mathematical abilities open up new possibilities and questions in the cognitive history of mathematics. Reciprocally, studies of mathematical cognition in context (c.f. Lave, 1988) applied to professional mathematicians at work (c.f. Livingston, 1986) can inform psychologists about what mathematicians really do. Such studies strongly constrain psychological theories about what and how mathematicians think. People adding large numbers use external artefacts as pen and papers rather than perform the prescribed algorithm mentally. Similar practices surely hold in the work of the professional mathematicians, and being aware of these practices could spare the psychologists from erroneously psychologising mathematical procedures. Mathematical thinking is best reflected in mathematics in the making; it cannot be reduced to the final institutionalised corpus of mathematical truths.

In the previous chapters, I have argued at length that the epidemiology of representations is a theoretical framework that questions both the social and psychological processes of knowledge production. The piece of epistemology of mathematics I have provided in this section shows that the history of mathematics would benefit of studies integrating social and psychological concerns. The consequence is that the epidemiology of representations provides a fruitful framework for the history of mathematics.

7.2 The number sense as a psychological factor of attraction towards the Newtonian calculus

Applying the epidemiology of representations to the history of mathematics implies focusing on the distribution of mathematical representations. It requires questioning why and how mathematical representations are distributed as they are, and paying attention to the psychological and social components of the mechanisms of distribution. Mathematical abilities can then be part of the account of a social and historical phenomenon: the evolution of mathematics. In particular, mathematical abilities can determine the content of mathematics by constraining which mathematical representations will be found convincing and appealing, which representations will more probably arise in mathematicians' mind and be used in their production of mathematical public representations. Mathematical abilities can be psychological factors of attraction in the cultural production of mathematics.

I will argue that what Dehaene (1999) has called the 'number sense' — a cognitive ability for representing, and thinking with, numbers — has been a psychological factor of attraction towards the notion of limit, when, at the beginning of the 18^{th} century France, mathematicians were striving to develop coherent notions for the calculus on the basis of the competing works of Leibniz and Newton.

7.2.1 The number sense as a cognitive ability – brief review of the psychological literature

The 'number sense' is a cognitive ability that has been much studied these recent years: evidence about its existence and functioning are found in the wide array of the cognitive sciences, from cognitive ethology to neuroscience (for comprehensive reviews, see Dehaene 1999; Gallistel and Gelman 2005). The number sense is defined as the capacity to quickly understand, approximate and manipulate numerical quantities. Dehaene (1999) argues that there are cerebral circuits that have evolved specifically for the purpose of representing basic arithmetical knowledge. The core of the cognitive theory is that humans and other animals are endowed with a mental system of representations of magnitudes, which is represent both continuous and discrete quantities. Humans use this representational system of magnitudes to comprehend number terms and do approximate calculation—This gives the number sense.

The arithmetical performances of animals and young babies constitute strong evidence of the existence of an evolved ability for representing and manipulating quantities. Experiments with pigeons, rats and monkeys as subjects have consistently shown their ability to evaluate quantities. Their performances go from ordering quantities to addition and subtraction, but also division and multiplication. Here are some instances of the experimental results:

Numerical competence in non-human animals (see Dehaene et al., 1998; Brannon and Roitman, 2003, chapter 1): Birds can consistently pick up boxes with the same number of spots on it —independently of the size, colour, and arrangement of the spots. The results of the experiments suggest that birds can count up to around six and well approximate larger numbers. Macaque monkeys are able to choose the larger of two sets of food items and lions are able to estimate whether their group is more numerous than another group, which shows that these animals can order quantities. Some vertebrates can also divide mumerosity by duration and obtain rates: when they are free to forage in two different nearby locations, moving back and forth repeatedly between them, the ratio of their stays in the two locations matches the ratio of the number of rewards obtained from unit of time. Moreover, experiments have shown that this equivalence must be calculated, because it is obtained much earlier than what would be possible with a trial and error cognitive strategy. Similar results show that some animals are able to multiply rate by magnitude.

Human non-verbal arithmetic abilities (see Lipton and Spelke, 2003; Wynn, 1998; Pica et al., 2004; Dehaene, 1999, chapters 2 and 3): very young babies have been shown to do mental addition and substraction. Babies, indeed, are able to anticipate the number of items that a box contains when (1) shown how many items are initially in the box, (2) shown items being added in or taken out the box. Babies are surprised (they look significantly longer) when the eventual number of items in the box is not the number that results from adding and substracting the number of items added or taken out. Developmental psychologists conclude that babies performed mentally the arithmetical operation, which when compared with the actual observed result lead to being surprised or not. Also, experiments in cross-cultural psychology, especially with people whose language does not include terms for numbers larger than five, strongly suggest that the ability to understand quantities and perform exact arithmetic with small numbers, and approximate calculation for large numbers, is universal across cultures.

Implementation of arithmetic abilities in the brain (Noel 2001; Dehaene 1999, chapters 7 and 8) :Evidence for the domain specificity of the ability to reason arithmetically has also been found in neuropsychology: there are selective preservation of arithmetical skills in the context of severe cognitive impairments such as semantic dementia (impairments of the ability to understand the meanings of words), and there are selective impairment of arithmetical skills. This gives people who are totally normal, but unable to say which from 28 and 99 is bigger; and people who are unable to name a fork but can calculate normally 13 times 25. Experiments with neuroimaging have also enabled localising brain areas which seem necessary for cognising quantities: the parietal lobes of the brain are involved in numerical cognition.

Psychologists have developed a model of the mental processes, which accounts for numerical cognition. This model first maintains that quantities are represented with mental magnitudes. There is a representational system that is put at work for understanding both uncountable quantities, as temporal magnitudes, and countable quantities, as number of dots or items of food. Mental arithmetic operations require processing mental representations. Also, magnitudes and numbers seem to be represented by the same type of mental representations. This is because, first,countable and uncountable quantities can be arguments of a single arithemetic mental operation, as when temporal magnitude is divided by the number of preys obtained. Second, a similar 'scalar variability' is observed when subjects manipulate magnitudes and numbers; where scalar variability characterises the fact that the larger is the quantity memorised, the less precise are the estimations of this quantity. A more specific phenomenon is Weber's law: the performance in discriminating two magnitudes is a function of their ratio. It is consequently asserted that the mental representations with which animals (human and non-human) understand quantities and perform basic arithmetic are mental magnitudes [3].

> Mental magnitude refers to an inferred (but, one supposes, potentially observable and measurable) entity in the head that represents either numerosity (for example, the number of or-

[3] This model of numerical cognition is adopted by most cognitive scientists working on arithmetical abilities, esp. Dehaene, Gallistel, Gelman, Wynn. An alternative account of numerical reasoning abilities put object-tracking capacities as the capacities from which numerical reasoning derive — see Wynn 1998 for a criticism of this view

> anges in a case) or another magnitude (for examples, the length, width, height and weight of the case) and that has the formal properties of a real number. Gallistel and Gelman (2005)

The formal properties referred to are: (1) for every line segment there is a unique real number that correspond to its length and conversely, for every real number there is a line segment whose length is that real number; and (2) the system is closed under its combinatorial operations (addition, subtraction, division, etc.): when applied to real numbers, these operations generate another real number. These properties are said to hold for mental magnitudes. Experimental evidences such as those mentioned above suggest that humans possess a domain specific biologically determined mental device which performs arithmetic operation on mental magnitudes.

The mental representational system is more akin to the real numbers than to the natural numbers, but how to account for our specific and privileged understanding of natural numbers? There is, first of all, the fact that we know at once how many entities there are, when they are less than six. This ability is called 'subitizing' and is presumably based on an object tracking system. Carey (2001) hypothesises that basic arithmetic (e.g. the successor principle) is first learned on the basis of operating with sets with number of elements within the range of the object tracking system. In any case, this ability can account only for the ability to represent small quantities. The accumulator model explains how the representational medium of mental magnitudes is used in counting:

> At each count, the brain increments a quantity, an operation formally equivalent to pouring a cup into a graduate. The final magnitude (the contents of the graduate at the conclusion of the count) is stored in memory, where it represents the numerosity of the counted set. Memory is noisy [...], which is to say that the values read from memory on different occasions vary [manifesting scalar variability]. (Gallistel and Gelman, 2005)

The system representing numerosities by mental magnitudes is homologous in preverbal children and non-verbal animals. With language, however, numerical cognition recruits further processes. Brain imaging furnishes evidences that linguistically based arithmetic cognition and non-linguistic arithmetic cognition are activating two distinct brain circuits. The psychological hypothesis about linguistically based arithmetical cognition is that it uses both the manipulation of symbols for calculation and the explicit algorithms taught (starting with verbal counting) *and* the mental magnitude representational system. Learning verbal counting and the ensuing numerical cognition therefore implies that:

> in the course of ordinary development, humans learn a bidirectional mapping between the mental magnitudes that represent numerosity and the words and numerals that represent numerosity. They make use of this bidirectional mapping in talking about number and the effects of combinatorial operations with numbers. There is broad agreement on this conclusion within the literature on numerical cognition, because of the abundant evidence for Weber-law characteristics in symbolic numerical behavior. The literature on the deficits in numerical reasoning seen in brain injured patients is broadly consistent with this same conclusion. It also seems plausible that the nonverbal system of numerical reasoning mediates verbally expressed numerical reasoning. (Gallistel and Gelman, 2005)

Skills in verbal counting are therefore combined with processes on mental magnitudes. Symbolically based arithmetic cognition is much slower than non-verbal arithmetical cognition, but issues exact, rather than approximate results. For instance, a set of thirteen items can either be counted — each item is named with a number term, the number term of the last named item is the cardinal of the set — or approximated — a mental magnitude is given to the set and then translated into a number term. Counting implies a long reaction time that increases as the set becomes larger; the reaction time even depends on the length of the number terms: French speaking people will take more time than Cantonese speaking people. By contrast, approximation implies relying on our number sense, the precision of the approximation decrease as the number increase but the reaction time remains very low.

The literature on naïve arithmetic and the the number sense says little about our understanding of the concept of limits or infinitesimals. Lakoff and Nuñez (2000), however, have hypothesised that the understanding of mathematical infinity relies on conceptual metaphors that use our conceptualisation of action (the "aspectual system"). Actual infinity is conceptualised as the result of iterative action that do not end. Lakoff and Nunez call this metaphor the Basic Metaphor of Infinity (BMI). While Gallistel and Gelman (2000; 2005) assert that the concept of real number is already present in our minds, Lakoff and Nuñez (2000) insists on the contrary that it results from metaphorical thinking. To begin with, real numbers importantly rely on the concept of infinity, which is understood with the BMI. The real numbers, indeed, include numbers with infinite decimals, solutions to infinite polynomials, limits of infinite sequence, etc. As for the Real line — the assertion that the reals are points on a line — Lakoff and Nuñez (2000) wittingly point out that our naïve understanding of a line need not imply that it is exhausted by the real numbers (i.e. there is a one to one mapping between the real numbers and the points of the line): a line can also be formalised with the hyperreals, with the consequence that the real numbers are relatively sparse among the hyperreals on that line. They also pin down the complex reasonings in the course of the history of mathematics through which the "the naturally continuous space" was thought in terms of discrete entities. They show that the real line is not directly derived from a naïve understanding of continuity, but is based on thinking of continuity as numerical completeness — a step initiated by Dedekind in 1872.

One cannot see more than important similarities between the mathematical, historically constructed, notion of real number and the mental system for representing quantities. For instance, we cannot really say that transcendental numbers have, as such, a corresponding intuitive mental representations. What Gallistel and Gelman have insisted on, rather, is that the mental system for representing quantities is more similar to the real numbers than to the natural numbers. The theory of the real number was *motivated* by the existence of mental representations of magnitudes that could not be expressed in the language of mathematics. For instance, we can have a mental representation of $\sqrt{2}$ as the length of the diagonal of a square whose sides are of length 1, although the rational numbers do not include such a representation. I think the existence of this cognitive motivation is the best way to understand Gallistel et al.'s assertion about the "platonistic rediscovery" of the real number. This motivation, however, under-determines the particularities of the real numbers as a mathematical construction. The existence of non-standard analysis can indeed be taken as a proof that the mathematics of quantities can evolve in many different ways. Why, indeed, shall we leave out Robinson's hyperreal numbers out of the "platonistic rediscovery"? In this condition, we are either led to say that all of mathematics is platonistic rediscovery, which is just restating the epistemically empty platonistic philosophy of mathematics, or we stay at the more modest claim that the mental lexicon for quantities is larger than the public lexicon furnished by the integer terms (see Sperber and Wilson 1998 for an argument that the mental lexicon is, in general, larger than the public lexicon) . A third solution is to say that the number sense has actually played a role in the history of mathematics, which favoured the construction of the real numbers. The hypothesis is then that the mathematical theorisation of the real numbers has been constrained by the pre-existing structure of our representations.

7.2.2 The two competing cognitive practices of the Calculus: Leibniz and Robinson versus Newton and Weierstrass

One important event in the history of the theorisation of the real numbers is the advent of the calculus: during a century the ontology of numbers was uncertain; the main question being whether infinitesimals were or were not numbers. The answer was eventually given in the negative:

the need for the notion of the infinitesimals, present as a methodological notion in calculations of derivatives and integrals, was eventually eradicated and replaced by a process: going to the limit. The calculus arose with the will to arithmetise phenomena observed in geometry — especially the existence of tangents to curves and the existence of surfaces delimited by curves — and in mechanics — such as the changing speed of falling objects. Two significantly different theories, one by Newton and the other one by Leibniz, were developed in order to effectuate this arithmetisation. Although the two methods lead to similar calculations and results, they are different at least because Leibniz method introduces new entities with which arithmetical operations could be done, the infinitesimals, while Newton did not appeal to infinitesimals but relied on a process where quantities are 'disappearing.' Infinity is present in both Newton and Leibniz's work, but it is present either in a new mathematical operation or in a new mathematical entity. These two approaches to the calculus played a competing role in the practice of mathematics during the 18$^{\text{th}}$ century and the first half of the 19$^{\text{th}}$ century. Guicciardini (1994) describes the 'cohabitation' of these methods as follow:

> During a very long and fruitful period, beginning with Isaac Newton and Gottfried Wilhelm Leibniz and continuing at least as far as Augustin Louis Cauchy and Karl Weierstrass, the calculus was approached and developed in several different ways, and there was debate among mathematicians about its nature. We can identify several different traditions before the time of Cauchy; one approach is to concentrate on three 'schools': the Newtonian, the Leibnizian and the Lagrangian.

Leaving out the less significant Lagrangian school, he says:

> The Leibnizian (mainly Continentals) and the Newtonians (mainly British) agreed on results — their algorithms were in fact equivalent — but differed over methodological questions. In some case this confrontation was influenced by chauvinistic feelings,

> and a quarrel between Newton and Leibniz and their followers, over the priority in the invention of the calculus, soured the relationships between the two schools.

Leibniz infinitesimal calculus is based on the idea that the mathematician can choose infinitesimal quantity and use them for calculation. Newton's fluxionary calculus aims to formally represent change through the geometrisation of time. Guicciardini (2003) gives a balanced account of the differences between the Leibnizian and Newtonian calculi:

> In my opinion, Leibniz's and Newton's calculi have sometimes been contrasted too sharply. For instance, it has been said that in the Newtonian version variable quantities are seen as varying continuously in time, while in the Leibnizian version they are conceived as ranging over a sequence of infinitely close values (Bos 1980, 92). It has also been said that in the fluxional calculus, "time", and in general kinematical concepts such as "fluent" and "velocity", play a role which is not accorded to them in differential calculus. It is often said that geometrical quantities are seen in a different way by Leibniz and Newton. For instance, for Leibniz a curve is conceived as polygonal — with an infinite number of infinitesimal sides — while for Newton curves are smooth (Bertoloni Meli 1993a, 61–73).
>
> These sharp distinctions, which certainly help us to capture part of the truth, are made possible only by simplifying the two calculi. As a matter of fact, they are more applicable to a comparison between the simplified version of the Leibnizian and the Newtonian calculi codified in textbooks such as l'Hôpital's *Analyse des infiniments petits* (1696) and Simpson's *The Doctrine and Application of Fluxions* (1750) rather than to a comparison between Newton and Leibniz.

In the next sections of this chapter, I will analyse why l'Hôpital's *Analyse des infiniments petits* has developed more radical views of the infinitesimal calculus, and I will attempt to explain why this radical view did not stabilise, but drew towards dualist methods in the calculus — using infinitesimals or evanescent quantities when needed — and then to the notion of limit. The publication of Cauchy's *Cours d'Analyse*, in 1821, is a key event in the evolution of the mathematical foundations of the calculus. It includes definitions of limits, continuity and convergence. Lakatos (1978), however, argues that Cauchy was still very much in the tradition of the Leibnizian calculus, relying on infinitesimals in the calculus. Lakatos shows that Cauchy's mistaken proof that the limit of a series of continuous functions is continuous is mistaken especially when anachronistically interpreted in the light of Weierstrass' theory. It is, indeed, only with Weierstrass' work, published in 1856 and after, that Leibnizian calculus and its reliance on infinitesimal quantities was abandoned. Weierstrass, then, ended dualist methods by imposing the notion of limit. This notion is the heir of the Newtonian notion of evanescent quantities, and it eradicates the notion of infinitesimals. It could therefore be said that Newton's ideas eventually won over Leibniz's. This phrasing is, of course, an oversimplification: ideas have evolved and transformed during the 18$^{\text{th}}$ century. But it expresses the fact that the underlying understanding of integration and differentiation is similar in Newton's formulation and in contemporary analysis. This similarity is all the more apparent because non-standard analysis, whose development began in the 1940', is, by contrast, more similar to the ideas of Leibniz than to the ideas of Newton. Indeed, the development of non-standard analysis, especially by Robinson in the 1960', has provided some new grounds, and mathematical honourability, to the notion of infinitesimals. Non-standard analysis is understood by those who developed it as a revival of the use of infinitesimals.

Why did the calculus evolve as it did? A teleological history of the calculus would assume that the concept of limit is the eventual, long waited for, discovery of the foundations of the calculus. The concept of infinitesimal was not a genuine mathematical notion — unclear as it was — and was therefore bound to disappear. Yet, non-standard analysis provides alternative rigorous foundations to the concept of infinitesimals, it shows that there exists of a set *R that contains both the real and infinitesimal numbers. The historiographical lesson of this historical event is that there must be some historical causes why the concept of infinitesimals was discarded for a century (say from Weierstrass work in the 1860′ to Robinson's work in the 1960′): the downfall of infinitesimals is not the mere and straightforward consequence of the lack of rigorous foundations. The concept of infinitesimals did stabilise during a century and a half, but it was always challenged by the ideas of evanescent quantities and going to the limit. Why did infinitesimals obtain some success while Newton did without? What caused the eventual downfall of Leibniz's theory before its renewal with non-standard analysis? Lakatos (1978) suggests an explanation: "it was the heuristic potential of growth — and explanatory power — of Weierstrass's theory that brought about the downfall of infinitesimals." Lakatos' idea is that the notion of infinitesimals, without the further mathematical theories that enabled Robinson to develop non-standard analysis, would not lead to 'refutable assertions' (where, in an application of Popper's theory of science to Mathematics, the content of mathematical theories is made of such assertions, and where refutability increases with the advent of rigorous proofs). With the theory of limits, the infinitesimals lost their power to bring about new results in the calculus; the same results could be found without appealing to infinitely small quantities. One can feel the blade of Occam razor in Lakatos' historical account. The two notions were redundant, so one of them could be eliminated at no cost, but why one notion was chosen rather than the other? As non-standard analysis shows, it is possible to do without the concept of limit and with the concept of infinitesimals, rather than without the concept of infinitesimals and with the concept of limit as in standard analysis. Occam razor could have eliminated Newton's evanescent quantities rather than Leibniz infinitely small quantities. In fact, one observes that the preference for the process the evanescence of quantities or going to the limit has right from the beginning undermined the appeal to infinitesimals (see next section). The preference for the notion of limit is also shown by the fact that the concept was independently discovered at different times and place: well before Cauchy's and Weierstrass' publications, Bolzano, in the Prague of 1817, published a satisfyingly rigorous definition of a limit (the epsilon-delta technique). This work remained unknown to the French and German mathematicians, so Bolzano's work cannot be said to have determined the thoughts of Cauchy and Weierstrass.

We are therefore in a case where:

1. Given two mathematical notions, one of them was privileged at the expense of the other.
2. There was a natural tendency to develop the notion of limit — as is most manifest with the case of independently enounced but similar definitions.

7.2.3 Why thinking with limits has been more appealing than thinking with infinitesimals

The above two characterisation of the evolution of the calculus suggest that there exists a *cultural attractor* towards the notion of limit. Furthermore, I will argue that the attraction towards the notion of limit is largely due to the way we naturally think about quantities, i.e. to the existence of the number sense. In other word, the number sense has been a psychological factor of attraction towards the notion of limit.

The two concurring models of Newton's fluxion and Leibniz's infinitesimals are based on different metaphors, thought processes and intuitions. Kurz and Tweney (1998), for instance, characterise thinking with Leibniz's calculus as thinking of oneself as the agent choosing infinitely small differences. By contrast, thinking with Newton's calculus involves transforming change into the continuous motion of a point on a graph. According to Lakoff and Nuñez (2000) both models use metaphors which eventually call on the Basic Metaphor of Infinity, i.e. taking the result of an unending process. Lakoff and Nuñez (2000) characterise the work of Weierstrass as taking part of the "discretization of the continuous." This programme in mathematics includes the Cartesian metaphor, where numbers are points on a line, and is further realised by the conceptual blend of the domains of space, sets and numbers, which especially took place in the 19th century.

In some ways, Lakoff and Nunez's account echoes the speculative hints of Gallistel et al. (2005): the history of Mathematics includes an appropriation, with mathematical symbols, of the naive perception of the continuous. Lakoff and Nunez peer more deeply into the content of mathematics than Gallistel et al. (2005); but Gallistel and his colleagues provide much more experimental evidence of their psychological hypotheses. Although they are opposed in several ways, it is worth using both works to shed light on the history of the calculus.

According to Lakoff and Nuñez (2000):

> [Weierstrass'] work was pivotal in getting the following collection of metaphors accepted as the norm:
>
> Spaces are set of points
>
> Points on a line are numbers
>
> Points in a n-dimensional space are n-tuples of numbers
>
> Functions are ordered pairs of numbers
>
> Continuity for a line is numerical gaplessness
>
> Continuity for a function is preservation of closeness

One important feature of this assertion is that the advent of these "metaphors", as constitutive of mathematical thinking, was not determined only by the properties of the human mind. The human mind could have used different metaphors for developing mathematics. In particular, the metaphors are used because they are furthering a research programme: the discretisation of the continuous (for an analysis of the role of research programme in mathematical practice see van Bendegem and van Kerkhove 2004; Kitcher 1984, chap. 7). This research programme is itself contingent on Mathematicians' interests: they especially wanted (and still want) to do away with thoughts based on drawings, judged approximate. Digital symbols were and are trusted as good means for mathematical reasoning, but analogical graphs were less and less trusted. The infinitesimal calculus is part of this travel from geometry to arithmetic, and is, in that respect, in the continuation of Descartes' analytical geometry. Thus Leibniz, in a letter to Huyghens (29 décembre 1691), writes "Ce que j'aime le plus dans ce calcul, c'est qu'il nous donne le même avantage sur les anciens dans la géométrie d'Archimède, que Viète et Descartes dans la géométrie d'Euclide ou d'Apollonius, en nous dispensant de travailler avec l'imagination [4]."

Weierstrass' definitions are now standards. They are:

Definition of the concept of limit

Let f be a function defined on an open interval containing a, except possibly a itself, and let L be a real number, then

$$\lim_{x \to a} f(x) = L$$

if and only if for all $\epsilon > 0$, there exist a $\delta > 0$ such that if $0 < x - a < \delta$, then $f(x) - L < \epsilon$.

Note that this definition is based on simple intuitions about comparing magnitudes — something that is straightforwardly done with the number sense (of course, understanding the definition also requires understanding the notion of function, the uses of the symbols, etc). The definition of derivatives is then based on the notion of limit:

$$f'(x) = \lim_{\epsilon \to 0} \frac{f(x + \epsilon) - f(x)}{\epsilon}$$

The above sentences do not directly contradict our intuitions about quantity. On the contrary, if continuous and discrete quantities are indeed intuitively represented with the same representational system, then they should be easily intuited. Dedekind (a contemporary to Weirstrass who defined the real numbers) is explicit about his goal when contributing to the calculus: maintaining arithmetic intuitions and applying them to the realm of the continuous. In his *Continuity and Irrational Numbers* (1872) he

[4] what I like most in this calculus, is that it gives us the same advantage over the ancients in Archimede's geometry, as Viète and Descartes in the geometry of Euclid or Appollonius, by dispensing us to work with imagination (my translation).

exposes his project of understanding continuity on the basis of the natural numbers, to which arithmetic applies. Considering geometric intuitions, he expresses his intention to do without them "in order to avoid even the appearance as if arithmetic were in need of ideas foreign to it" (p. 5, quoted in Lakoff and Nuñez 2000, p. 295). Of course, the definitions of limit and derivatives are distinct from Newton's definition: they involve static relations among points while Newton appealed to movement. However, they are similar to the extent that they link geometrical intuitions to arithmetic intuitions, thus bringing the inferential power of the latter to understand better the former. Going to the limit is a process that still calls on the number sense, while infinitesimals are entities that contradict intuitions provided by the number sense. Take, for instance, L'Hopital's "demand" at the beginning of his *Analyse des Infiniments Petits* (1696):

> 1. Demande ou supposition. On demande qu'on puisse prendre indifféremment l'une pour l'autre deux quantités qui ne différent entr' elle que d'une quantité infiniment petite : ou (ce qui est la même chose) qu'une quantité infiniment moindre qu'elle, puisse être considérée comme demeurant la même. [5].

This postulate is thus saying that $x + dx = x$, where dx is a quantity that is infinitely smaller than x. L'hopital presented this postulate as something obvious, both in conformity with our intuitions and already present, if not formulated, in the work of past mathematicians.

> D'ailleurs les deux demandes ou suppositions que j'ai faites au commencement de ce Traité, et sur lesquelles seules il est appuyé, me paroissent si évidentes, que je ne crois pas qu'elle puissant laisser aucun doute dans l'esprit des Lecteurs attentifs. Je les aurois même pû démontrer facilement à la manière des Anciens, si je ne me fusse propose d'être court sur les choses

[5] Demand or supposition [postulate]: we demand that it be possible to take indifferently one or the other of two quantities that differ only by a quantity that is infinitely small; or, (which is the same thing) that a quantity to which one add or subtract a quantity that is infinitely lesser than it, can be considered as remaining the same (my translation)

> qui sont déjà connues, et de m'attacher principalement à celles qui sont nouvelles. [6].

L'Hopital's confidence in the intuitive appeal of his postulates is not mere wishful thinking. There are some intuitions upon which one can base thoughts with infinitesimals. Common images such as the dune and the grain of sand metaphor can be called on for furthering understanding. Also, the postulate can indeed be presented as a valid interpretation of the Ancient's work (by which it is supposedly meant Archimedes' writings on the method of exhaustion, Cavalieri's *Geometria indivisibilibus* (1635), Roberval's *Traite des indivisibles* that introduces infinitesimal quantities in the calculation of surfaces and volumes, Fermat's procedure which uses the new analytical geometry). The point I want to make, anyhow, is that this intuitive ground cannot be the number sense. The representational system of magnitude does not include different scales for order of magnitudes that are incommensurable, i.e. it does not include different sets of representations of magnitudes across which addition or subtraction does not increase or decrease the initial amount.

Do we have an näive understanding of incommensurable magnitudes? I submit that we most probably do not. There is one single representational system for magnitudes across which arithmetic operations uniformely apply. When communicating, words such as 'small' can appeal to different scales; for instance when we use 'small' to qualify a small elephant and a small mouse. These linguistic facts are compatible with the hypothesis that there is one single mental representational system for quantity: ranges of possible size are pragmatically inferred and expressed within the representational system; there is no need to appeal to incommensurable mental magnitudes. It is also said that our visual representational repertoire is made of "middle-size objects". For instance, in order to represent things

[6]In passing, the two demands or suppositions that I have made at the beginning of this treatise [the demand one above quoted and a demand that concerns the definition of a curve] and upon which it is entirely based, appear to me so obvious, that I do not think they could leave any doubts in the mind of careful readers. I could even have easily proved them in the fashion of the Ancients, if I had not had the goal of being brief on those things that are already well known, and principally work on new ones (my translation).

that are very small such as atoms, we represent them as middle size objects, then add the further assertion that they are *not* at the size we may represent them, but infinitely smaller. There are two representations to obtain the final understanding of magnitude: a representation of magnitude directly derived from some public representation of the atom, and a representation evaluating to which extent the previous representation is signifying the actual magnitude. In either physics or mathematics, representations of infinitely small magnitudes have been produced through long histories of theoretical developments. Also, the psychological literature on the number sense seems to assume that we do not have intuitive representations of infinitesimal quantities. Unfortunately, I am not aware of psychological experiments *directly* tackling the question: the work of Gallistel, Gellman, Dehaene, and their collaborators does argue that we have mental representations of magnitudes that correspond to irrational quantities (such as $\sqrt{2}$), but it says little about *not* having infinitesimals included in the mental representational system he has been studying. One important property that distinguishes the reals from the hyperreals, which include the reals and the infinitesimals, is the Archimedean property. Having the Archimedean property means that:

$$\forall x > 0, \forall y, \exists n, \text{a natural number, such that } n.x \geq y$$

Does the representational system of magnitude have this property? Is the number sense Archimedean? Experimental evidence in favour of a positive answer would certainly corroborate Gallistel's assertion about the privileged relation between the real numbers and mental representations of quantities. Accepting infinitesimal quantities imply renouncing to the Archimedean property, since there is no natural number n such that $n.dx \geq x$.

Newtonian and Leibnizian calculi stand on different metaphors, intuitions and thought processes. Among the intuition used, Newtonian calculus keeps and uses the number sense, while Leibnizian calculus seems to limit its inferential power. The number sense is straightforwardly available when doing arithmetic calculation: as we have seen, learning to count and do arithmetic involves activating the number sense. Relevance theory and the epidemiology of representation tell us that models that allow theoretical statements to take a grip on our intuitions are preferred. This is because theoretical statements that have a grip on our intuitions enable intuitive inferences; they have relatively higher cognitive effect for lower processing effort, and are therefore more relevant. I hypothesise that Newtonian calculus and the concept of limit trigger the number sense in such a way that the inferential potential of this ability is well exploited. By contrast, Leibnizian calculus uses quantities, the infinitesimals, which either do not fit the domain of the number sense, or go against the inferences that this cognitive device makes. Interpreted in the cognitive perspective where inferences are enabled by the recruitment of domain specific abilities (c.f. previous chapter), this means that infinitesimals could not lead to a rich production of intuitive beliefs through the activation of the number sense. They put the mathematicians in an uneasy position as to which inference to make with arithmetic operations. As already mentioned, Lakatos (1978) asserts that the concept of infinitesimals was rendered useless by Weierstrass's theory. He thus explains why one of the two concepts—of limit and of infinitesimals—had to disappear: one of them was made irrelevant since redundant. Yet, this does not explain why the concept of limit was chosen rather than the concept of infinitesimals. The explanation of the choice relies on a further psychological hypothesis: inferences based on naïve arithmetic are blocked in the infinitesimal calculus, thus requiring more effortful non-intuitive inferences to achieve the same cognitive effect. Actually, this loss of intuitively derived cognitive effect is largely compensated with the inferential potential of the calculus. On the whole, then, there is a gain in cognitive effect which explains why infinitesimals have had some cultural success in the 18$^{\text{th}}$ century. However the concept of a limit achieves the same increase in cognitive effect provided by the calculus without forsaking the inferential power of the number sense. It is therefore more relevant than the concept of infinitesimals. The concept of limit keeps arithmetic intuitions of the number sense, rely on the number sense, and at the same time achieve the goals set by the calculus. The concept of infinitesimals eventually re-entered mathematical knowledge when the new mathematical context gave it some supplementary cognitive effect. The number sense has acted, at the end of the 17$^{\text{th}}$ century, as a psychological factor of attraction, increasing the probability of distribution of representations similar to the notion of limit, which makes the most of the inferential power of the number sense.

Hypothesising the existence of a psychological factor of attraction is not teleologistic in the classical sense. The hypothesis asserts that given the state of mathematics at the time and the human cognitive capacities, then the *probability* that the calculus developed as it did was high. The teleological component of the hypothesis is justified by the specification of the causal processes that make it true. The assertion is not that Mathematics was bound to be what it is because of some unexplained necessity. Rather, the hypothesis points out that psychological processes are such that, in the specific cognitive environment of the time, a mathematical notion is more *appealing* than another one with similar function. From this, one deduces that the probability that the more appealing notion be taken on by the mathematical community is higher than the probability that the less appealing notion be taken on. At the social level, there is a process of distribution of representations that distributes with greater ease and probability representations that are more similar to the notion of limit than to the notion of infinitesimals. Getting down to the psychological details, the difference of appeal of the two notions is explained in terms of their respective relevance to the mathematicians of the period. The psychological hypothesis can be made sensitive to cultural changes: when non-standard analysis was developed in the mid-nineteenth century, the notion of infinitesimals had become appealing again. A last problem with the teleological aspect of the hypothesis is the anachronism it seems to be based on: why can we use the notion of limit in order to interpret two mathematical notions that predate it? Neither the notion of infinitely small quantities nor the notion of evanescent quantities tacitly includes the notion of limit. The latter, indeed, requires an understanding of the notion of function, which will appear only much later. So what does it mean that the notion of limit was a cultural attractor that favoured the distribution of Newtonian representations of evanescent quantities rather than the Leibnizian representations of infinitely small quantities? The reason why the notion of limit is helpful for understanding what is at stake at the psychological level is that Newton's notion and the notion of limit are based on similar thought processes, they use of the same underlying metaphors for understanding infinity.

The hypothesis about psychological factors of attraction is not psychologistic either. It is not assumed that the concept of limit is a psychological primitive; that it belongs, for instance, to the innate concepts of a language of thought. Attraction towards the notion of limit is not caused by the discovery of one's own underlying cognitive processes. It is not a process of externalisation, in public representations, of mental representations. The process of attraction relies on the differential relevance of competing notions. Thus the notions of limit and infinitesimals can still be considered as what they really are: historical conceptual constructions rather than concepts universal to the human species. And yet, some psychological reality does determine the history of the concepts. Most of the framing of the notions of the calculus in the 18^{th} and 19^{th} century had to do with the choice of a model that would enable achieving explicit goals (e.g. calculating surfaces delimited by curved lines) at the minimal expense of arithmetic intuitions (esp. naïve arithmetic).

7.3 MECHANISMS OF DISTRIBUTION OF MATHEMATICAL REPRESENTATIONS

L'Hopital's first axiom in his *Analyse des Infiniments Petits* (1696), the equation $x + dx = x$, could not be taken for granted. A lot of background knowledge was brought up to show the relevance of making such an assumption: this included the goals of calculating surfaces and rates and the previous means developed to satisfy these goals, such as the method of exhaustion [7]. It is only after two century and a half of calculus, from Leibniz to Robinson, that mathematicians have come to think of infinitesimals with sufficient ease. As is well known, the history of the irrational numbers has known a similar fate, and it lasted much longer to get from the discovery of irrational quantities to an ease of use of these quantities in Mathematics. An epidemiological analysis could possibly show that these histories differ nonetheless in their appeal to intuitions. The epidemiological rendering of Gallistel et al.'s (2005) hypothesis is that a driving force in the mathematical theorisation of the real number line was the existence of our mental system for representing magnitudes — mathematicians had a mental representation of the length of the diagonal of square of side one, but could not, at first, do mathematical reasoning with this representation. The epidemiological hypothesis with regard to the evolution of knowledge about infinitesimals is that it was blocked by a negative difference of relevance with the concept of limit. In the following, I analyse the mechanisms that contributed and hindered the distribution of the notion of infinitesimal. I begin by emphasising factors of distribution that are to a certain extent independent of the content of the notion distributed, then I point out where psychological factors of distribution may have intervened in the evolution of the concept of limit. My analysis is essentially based on second sources history, especially the accounts of Boyer (1959); Robinet (1960); Blay (1986); Mancosu (1989); Jahnke (2003).

7.3.1 Trust-based mechanisms of distribution: Malebranche as a catalyst

The concept of infinitesimals, as many mathematical concepts, did not stem from an individual mind at a precise time in history with a precise and definitive meaning. It has a history during which its future use was being determined. The concept of infinitesimals travelled through

[7] Classical milestones before the introduction of the calculus in France are Archimedes' writings on the method of exhaustion, Cavalieri's *Geometria indivisibilibus* (1635), Roberval's *Traite des indivisibles* that introduces infinitesimal quantities in the calculation of surfaces and volumes, Fermat's procedure which uses the new analytical geometry and eventually Leibnitz'*Meditatio Nova* (1686).

time- its history can be traced to Zeno's paradoxes (Vth centery B.C); but also through disciplines- from theology [8], natural philosophy (mechanics) and geometry to arithmetic; and through schools of thought—such as from Leibniz's formalism to Malebranche's initial Cartesianism. Of course, mathematical concepts do not travel by themselves; they travel because of people's action. The analysis of the history of mathematical concepts is therefore an analysis of mathematicians' actions and thoughts. The notion of infinitesimals first travelled from Saxe to France through Leibniz' correspondence with Malebranche. The actual introduction of the calculus in France is due to J. Bernouilli's visit to Paris. When he arrived, in 1691, he went directly to Malebranche. This move was decisive, for he met in Malebranche's room the Marquis de L'Hopital, to whom he taught the calculus during the winter 1691-1692. The result of this tuition is the book *Analyse des infiniments petits*, which remained the French reference book in the calculus for a century. In all these events, Malebranche played an essential role. He was a catalyst in the process through which French mathematicians came to study the calculus. One can distinguish two stages in the process of distribution of the calculus: the first stage is when mathematicians get to know the calculus, the second stage is when they become convinced of its worthiness and actually use it and work with it. Malebranche proved indispensable at both stages. Although the calculus was available to French mathematicians as early as 1684, with Leibniz' "Nova Methodus", it was only after Leibniz personally convinced Malebranche of the importance of the calculus that contemporary mathematicians began to consider this new theory. Malebranche was a European figure and a promoter of sciences. This, together with his interest in mathematics, made him both the link between the source of the calculus, Leibniz, and the French mathematicians, and the leader of the movement for the cal-

[8]The theological connotations of the concept of infinity is apparent until the 18th century. This can be seen for instance in Pascal reflexion on the mathematical operation with infinite quantities: "L'unité jointe à l'infini ne l'augmente de rien, non plus qu'un pied à une mesure infinie. Le fini ne s'anéantit en présence de l'infini, et devient un pur néant. Ainsi notre esprit devant Dieu; ainsi notre justice devant la justice divine. Il n'y a pas si grande proportion entre notre justice et celle de Dieu, qu'entre l'unité et l'infini." *Pensées*,f3, sect. III, fr. 233.

culus in France. He parted from Prestet and Catelan, his previous Cartesian mathematician disciples, and constituted around him a new group of mathematicians whom he directed toward the calculus. He also participated actively to the development of the calculus with criticisms and comments [9]. The introduction of the calculus in France is done through a process of distribution of representation that relies on the recognised epistemic authority of those that first used the representation. Boudon (1979) illustrates the process with Hagerstrand's study of the diffusion of an agricultural innovation in Sweden, which shows that the adoption of a new technique is a process that requires social actors' "confidence". This confidence can only be attained by being exposed to a "personal influence". Once this is achieved, the new technique spreads because of what Boudon calls the "imitative dimension" of social actions. The calculus was, in 1690, a new technique and one can recognise in Malebranche, and later in the Infinitesimalists, the personal influence necessary to its spread [10]. As in the case of the Swedish agricultural innovation, the existence of the new technique alone was not sufficient to overcome the "intrinsically convincing traditions" that were the Cartesian and synthetic practices of mathematics. The influence of epistemic authorities has been decisive in the progressive change of mind of the academicians and, later, that of the wider community of mathematicians. It explains the fact that the calculus was taken on by only a few mathematicians, and then accepted at an exponential rate (the more mathematicians there are, who have adopted the calculus, the more influence there is for convincing other mathematicians). Also, countries without their Malebranche did not develop interest in the calculus as in France. There is, in the process of distribution of scientific ideas a bias to imitate, or follow, those individuals that proved to be successful (Boyd and Richerson, 1985). Bloor (1996) mentions another important process of distribution of scientific ideas that is akin to the processes of adoption of technical innovation: once a technical standard or technology is adopted

[9] Volume 17 of Malebranche's *Oeuvres completes* contains Malebranche's encouragement and participation to L'Hopital's work.

[10] For instance, L'Hopital wrote to Malebranche that only his approval afforded him any satisfaction with his work (letter to Malebranche, 1690).

by a small but critical number of people, then the standard quickly spread and definitely prevail over competing standards or technologies. This is because once the technology is adopted by neighbours and friends, one will benefit in choosing the same technology because it opens up possibilities of cooperation. Likewise, once a technique or a theory is sufficiently well ingrained in the scientific practices, a scientist's has interests in using currently used techniques and theories so as to increase "possibilities for some form of cooperation, for example exploiting the work of others and making contribution of a kind that will be used and recognised." The recognition of the calculus as a mathematical theory can be characterised as a 'conquest' that the concept of infinitesimal made of the French Royal Académie des Sciences — once this conquest was made, the critical state of adoption was met and the calculus would impose itself on other mathematicians.

7.3.2 Interests and strategic means of distribution: aiming at the institutional recognition of the calculus

At the end of the 17th century the concept of the infinitesimal was not in accord with Cartesian principles. Its introduction in France therefore met strong opposition, which was concretised in the dispute that took place at the French Académie between the Infinitesimalists around Malebranche, and the Finitists. Malebranche received a group of mathematicians regularly in his room at the Oratoire, which then became the headquarters of the group. They developed so much interest for the calculus that, in 1699, the Malebranchists and the Infinitesimalists became one single group which struggled for the recognition of the calculus. The most active of them were L'Hopital, Varignon, who were already members before the reform of the Académie in1699, and Carre, Saurin and Guisnee who entered the Academie with Malebranche. The Infinitesimalists soon formed a compact group of interest that struggled for the recognition of the calculus. The recognition, largely due to Leibniz, that the calculus constituted an independent field, gave the Infinitesimalists a definite object to fight for. Another factor in their unity was the existence of an active opposition. The anti-Infinitesimalists, or finitists, had for champions Ph de la Hire, Galloys, and, chiefly, Rolle. The Académie was the greatest scientific French institution and was therefore worth conquering. The Malebranchists presented most of their work, as shown by the reports of the Académie, during the sessions of the scientific institution, and Malebranche himself assiduously attended them even before being appointed honorary member. The Academie was organised for the discussion of scientific problems. These made it the obvious site for the controversy that took place during the years 1700-1706. The main element of the controversy was an exchange of arguments between Rolle and Varignon. The debate, however, had essential political strategic components (Mehrtens (1994) argues that mathematics as any other science is bound to be political).

One important goal for the infinitesimalists was to win the approval of the scientific community at large. Varignon insisted to make the debate open to non-members of the Académie. Fontenelle's distinction between mathematical and metaphysical infinite (1727, p. 53) could be viewed as an attempt to reassure theologians and metaphysicians. Otherwise how would Cartesians, who derived from the idea of infinity the existence and nature of God, admit that the very same idea could be used to solve the brachistochrone problem? Another strategy used, was to insist on the ability of the calculus to solve problems and on the power of its methods, and to elude the problems of foundations. The control of means of communication was an important stake. Fontenelle, using the power that his position of secretary of the Academy conferred upon him, delivered in 1704, at the peak of the Infininitesimalists-finitists dispute at the Academie, a eulogy of L'Hopital in which he included a eulogy of the calculus. Another essential way to communicate one's ideas is through publishing, and so, a close relationship with the publishing trade was part of the strategies to acquire the approval of the community. The fact that the anti-Infinitesimalists

Gouye and Bignon were directors of the *Journal des savants*, the most important French scientific revue of the time, gave them an important advantage over the Infinitesimalists. Because of this, Varignon, writing to John Bernoulli, complained that the infinitemalists' answers to Rolle were being truncated when published in this Journal. But the Infinitesimalists were in control, via Fontenelle, of the report and of the official history of the Académie. The strategies with regard to written communication can again be seen in Malebranche's advertisement of L'Hopital's book, which replaces, in the 1700 edition of #, the one of Prestet's *Elements de Mathématiques*. These strategies eventually aimed at a favourable outcome of the debate over the calculus, the end of opposition to it, and hence its total recognition. Another stake was the organisation of the debate between finitists and infinitesimalists: the most obvious issue concerned the nomination of a commission to judge the dispute in the Académie, for this judgement would bring an official recognition or rejection of the calculus. In 1701 the anti-Infinitesimalist Abbe Bignon, then president of the Académie, nominated a commission composed of three people, two of whom were favourable to Rolle. Due to the increasing consensus on the calculus, this commission was unable to give an unfavourable judgement and postponed the decision until 1705 when a new commission, again favourable to Rolle, replaced it. In 1706 the new commission had to take into account the composition of forces within the Academie (predominantly infinitesimalists); thus Rolle was asked to stop the dispute. (c.f. Mancosu, 1989, pp. 239–40)

The introduction of the calculus in France was therefore partly the outcome of a dispute led by united groups of mathematicians. The success of the calculus is the result of the actions and strategies of the Infinitesimalists. As Mancosu (1989) says, "Mathematics and its development are due to human efforts and not only to the soundness of the ideas involved." Analysis of the human actions involved in the introduction of the calculus in France reveals them to be causes of the success of the calculus. Mathematicians as social actors succeeded in socially imposing the concept of infinitesimals as a genuine mathematical concept.

The strategies discussed above take place in a cultural setting and acquire efficiency by using cultural components. The victory of the calculus against what Varignon called the 'old style mathematicians' was partly due to the values of the time. These, used by infinitesimalists in they favour, enabled them to overcome the difficulties arising from the lack of rigor of calculation with infinitesimals. Infinitesimal quantities have no rigorous meaning (This is true both in today's and in the Cartesian 17$^{\text{th}}$ century's sense of mathematical rigor): This is Varignon's point, arguing that sometimes they were used as finite quantities, in equations of the type $(y.dx)/dx = y$, and sometime as zeros, such as in $x + dx = x$. But the calculus enabled to solve a tremendously wide number of problems, both mathematical and physical; this corresponded to the values of late 17th and 18th Century France. First, in the course of the scientific revolution it became apparent that mathematics could tell us something about the world — and indeed the calculus applies to mechanics; second, the utility of sciences, as shown by Fontenelle's preface to *L'Histoire de l'Academie Royale des Sciences* (1725), was sciences' best justification. Hence the calculus was developed notwithstanding its lack of rigour. The time of the 'siecle des Lumieres' had arrived and with it a new philosophy of mathematics in which analysis could grow. The cultural context of confidence in the progress of mathematics and its applications accounts for the success of the calculus and the outcome of the dispute which took place at the French Académie.

It is not just that socio-cultural components favoured the distribution of the concept of infinitesimal among the French scientific community. These components also determined the content of the concept. In the 17$^{\text{th}}$ century, the status of the infinitely small was problematic. While Leibniz sometimes gives it a purely formal status, the French infinitesimalists adopt a very realistic stance. Why do they do so, and what is the consequence for the evolution of the calculus? Describing how mathematical knowledge is evolving, Lakatos (1976) use the metaphor of a factious classroom and endows its pupils with different patterns of responses to unexpected Mathematical elements: these patterns include 'monster-barring'

— a knowledge strategy that consists in dismissing counter-examples to known theorems, maybe by re-specifying definitions — and 'exception-barring' — a strategy that consists in accommodating anomaly by drawing more subdivisions. Bloor (1978) further argues these patterns of responses may be determined by the social situation of the mathematicians or scientists. Following that trend, one can characterise the Infinitesimalists of the late seventeenth century as a small-threatened group. This explains the strategies they used in developing the knowledge of the calculus: they adopted a categorical stance which asserted the real existence of infinitely small quantities and strongly lamented Leibniz' hesitations with regard to the nature of those quantities. Their eagerness to go forward, showing more and more of the potential of the calculus, and the fact that they barely took the time, under the pressure of the finitists, to stop and think about the foundational problem, is a strategy of justification that can be compared to the strategy of the 'nouveaux riches' who, aspiring to the aristocrat status, display all their wealth. In the same vein, the Infinitesimalists' also called on previous well-known mathematicians to support their claim for recognition. Thus, Varignon asserts that "Mr de Fermat luy-même" used approximation. This is consequential on the evolution of scientific and mathematical knowledge. Indeed, the above strategies clearly influenced the practice and notions of the calculus. The realist philosophy towards infinitesimals allowed the bold development of equations with infinitesimal quantities. The legitimisation of their approach by reference to canonical works forced them to establish their continuity with tradition, and the emphasis on results granted the continuation of the development of the theory. The social context, that is the social values of efficiency, and the fight for recognition, induced the Infinitesimalists to make the calculus of the turn of the eighteenth century as it was: an aggressively assertive conqueror who, at the same time, was slowly framing his notions and rules.

An important means of targeted distribution of representations in the mathematical community implies 'Mathematising' terms. This implies showing the relevance of one's discourse to a relatively autonomous community, with its own goals and culture. Mathematising terms is achieved, in particular through the creation of symbols and by obtaining theoretical autonomy.

The concept of infinity was brought to mathematics from theology, philosophy and physics. The mathematical revolution of the calculus corresponds to the creation of a meaning for the concept of infinity that is proper to mathematics. “Le véritable continu est tout autre chose que celui des physiciens et celui des métaphysiciens [11] ” says Poincaré (1902). The process of emancipation of the concept of infinity from other disciplines includes the use of mathematical symbols, as those introduced by Leibniz. The introduction of symbols in mathematics has strong consequences on the cognitive practices in mathematics, and also on the specific meaning of the terms involved. As argued by Goody (1977), ”symbolic logic and algebra, let alone the calculus, are inconceivable without the prior existence of writing” p. 44. He further says:

> The increased consciousness of words and their order results from the opportunity to subject them to external visual inspection, a process that increase awareness of the possible ways of dividing the flow of speech as well as directing greater attention to the ’meaning’ of the words which can now be abstracted from that flow [. . .] The process is not simply of ’writing down’, of codifying what is already there. It is a question of formalising the oral forms and in doing so, changing them into something that is not simply an ’oral residue’ but a literary (or proto-literary) creation. (p. 115–6)

In the case of the infinitesimal the passage is from graphs to formuli, which led Leibniz to say that the calculus dispenses us to work with our ‘imagination.’ The new symbols introduced by Leibniz induced new ways to think with the concept of infinity. The symbols constrained in their own way how the concept was to be used. This has for consequences to give

[11] The true continuum is completely different from the one of physicists and metaphysicists (my translation)

autonomy to mathematical practices and to fix a specifically mathematical meaning to the notion of infinitesimals: as one specifies how to manipulate the symbol for infinitesimals, dx, one also specifies how the concept is to be used and understood, and theological or physical considerations are made much less relevant. Thus Poincaré (1902) says:

> L'esprit a la faculté de créer des symboles, et c'est ainsi qu'il a construit le continu mathématique, qui n'est qu'un système particulier de symboles. Sa puissance n'est limitée que par la nécessité d'éviter toute contradiction; mais l'esprit n'en use que si l'experience lui en fournit une raison [12].

The creation of an autonomous mathematical discourse is done through diverse means, which include denying the relevance of other discipline and the constitution of esoteric means of communication. For Cavailles (1938) the autonomy of mathematical discourse is an essential characteristic of Mathematics:

> Le mathématicien n'a pas besoin de connaître le passé, parce que c'est sa vocation de le refuser : dans la mesure où il ne se plie pas à ce qui semble aller de soi par le fait qu'il est, dans la mesure où il rejette l'autorité de la tradition, méconnaît un climat intellectuel, dans cette mesure seule il est mathématicien, c'est à dire révélateur de nécéssité [13].

This obviously contrasts with the infinitesimalists recurrent appeal to the "Ancients." This contrast is not, I believe, only due to possible change in the epistemology of mathematics, for in fact the infinitesimal calculus did consist in denying the methods of the ancients in order to replace it by new methods. Continuity and revolution is here a matter of degree. The

[12]The mind has the faculty to create symbols, and this is how it constructed the mathematical continuum, which is nothing but a particular system of symbols. Its power is limited only to the necessity to avoid any contradiction; but the mind uses it only when experience provides it with a reason to do so (my translation)

[13]The mathematician has no need to know the past, because it is his vocation to refuse it [. . .] to the extent that he rejects the authority of tradition, ignore an intellectual climate, to this extent only he his mathematician (my translation).

epistemological point of Cavailles applies to the infinitesimalists because the growth and recognition of their theory, as a mathematical theory, includes a process constitutive of autonomy. However, one sees that this process of acquiring autonomy is itself a social process. It implies the constitution of a group — the infinitesimalists in our case — with its specific goals and means.

7.3.3 An effect of psychological factors of attraction in the history of the calculus

The above epidemiological analysis shows how social interactions have favoured the infinitesimal calculus over the fluxional calculus. In France, the distribution of Leibniz's work was much wider than the distribution of Newton's work. It is only with the work of Maupertuis, Voltaire and the Marquise du Châtelet, in the second third of the 18$^{\text{th}}$ century, that the work of Newton was promoted in France (e.g. Voltaire's *Eléments de la philosophie de Newton* (1738) and the Marquise du Châtelet's *Institutions de Physique* (1740), followed by her translation of *Philosophia Naturalis Principia Mathematica*, from latin, in 1756). These authors have mostly defended Newton's theory of attraction against Cartesian physics. The work of Newton in mathematics was known much earlier on the continent, if only because of the priority dispute between Newton and Leibniz over the 'discovery of the calculus' (Newton and, with him, the Royal Society accused Leibniz of plagiarism). Yet, although Newton's work on the calculus dates back to the years 1665–1667, and although some results were published in his *Philosophia Naturalis Principia Mathematica* in 1687, it is only in 1704 that Newton published a systematic treatise on the calculus, called *De quadratura curvatum*, while Leibniz successfully promoted his work on the continent early on. His early publishing of *Nova Methodus* (1684) and *Meditatio Nova* (1686) in the newly created journal *Acta Eruditorum* (since 1682), his communications with Malebranche, the Bernouilli brothers and other mathematicians of the epoch, have been all successful means of distribution of his ideas. Contemporary standard analysis keeps much of Leibniz's view on the calculus — his symbols d and $\int$, most notably. Guicciardini (2003, p. 73) also says that the algorithm we employ today in solving differentials and integrals are more similar to Leibniz's than to Newton's algorithm's. And yet, the notion of limit is much more similar to the ideas of the Newtonian calculus than to the idea of infinitesimals developed by Leibniz and his followers. The striking fact is that Newton's notion of evanescent quantity appeared very early in French Mathematics — much before the work of Newton was well distributed in France. I still refer here to the dispute that took place at the Académie Royal between Varignon and Rolle from 1700 to 1706, where explicit appeal to Newton's idea was made by the advocate of the calculus.

Historians have pointed out that the exchange between Varignon and Rolle was not of great quality. Rolle's examples actually contained some mistakes in the proofs of his pseudo-counter-examples and Varignon's answer is qualified as 'puns' (Blay 1986, p. 232, Mancosu 1989, pp. 232–234). For Rolle, dx was not given the same meaning in the two equations $(y.dx)/dx = y$ and $x + dx = x$, while for the infinitesimalists dx is used in the same way, in accordance with its definition. A judgement of identity is being questioned among professional mathematicians. Such questions are important events in the development of science, because the answers provides the important 'exemplars' of how to use of the terms. Some authors in science studies would say that the meanings of scientific terms are being negotiated. The debate taking at the Académie des Sciences, is such a case where the meaning of mathematical terms is being specified. For Rolle, infinitesimals are monster numbers, as they do not comply with fundamental rules of arithmetic. He adopts the strategy that Lakatos calls 'monster-barring', attempting to deny the existence of the monsters. Varignon, by contrast, tries to reconcile arithmetic intuitions and the existence of infinitesimals.

Rolle's objection against the infinitesimal calculus as represented by l'Hôpital's *Analyse des Infiniments Petits* bore on the foundations of the calculus. The argument was that the infinitesimal calculus added nothing to the method of the Ancient — he especially refers to the method of Hudde — but lack of conceptual rigor and mistakes. In order to pin down a mistake made by the method of the calculus, Rolle had to show that, given a problem, the answer obtained by a secure method, namely Hudde's method, differed from the answer obtained by the infinitesimal calculus. Rolle's attempt on this point failed and it was shown that his presumed proofs of counter-examples included mistakes, or misuses of the calculus. Note that when Berkeley designed his own attack against the calculus, some thirty years later, he was quick to explain that his argument did not bear on the results, but on the rigor of the reasoning. It is also on the problem of rigor that Rolle's attack is to be taken seriously, and especially on the justification why the Archimedean property could not hold when working with infinitesimals. Why can we say, as in $x + dx = x$, that the part is equal to the whole? Varignon's answer is made striking by the fact that it draws on both Newton's and Leibniz's calculi. Mancosu (1989, p. 235) analyses Varignon's argument as follow:

> Varignon made use of Newton and Leibniz at the same time. Although Varignon espoused the Leibnizian formalism he interpreted the differential dx as a process, i.e., the process by which quantity x became zero (dx represented the instant in which x became zero) [...] in fact, dx functioned as a numerical constant, and, interpreting it as a process, Varignon's approach created an asymmetry, an incongruity, between the formalism and its referents.
>
> Varignon took for granted that the Leibnizian calculus and the Newtonian calculus were equivalent and that Newton's version was rigorous. This kind of assumption can be found later in the century.

We have seen that the infinitesimalists had a realistic stance for infinitesimals, while Leibniz himself took infinitesimals as well grounded formal entities ("on a pas besoin de prendre l'infini ici à la rigueur", Leibniz said). Together with this stance, the infinitesimalists still assumed that no new algebraic laws were needed for the infinitesimals. The way out of the problem was to give a dynamic interpretation of infinitesimals that drew on Newton's fluxion. Varignon had been working on application of the calculus to mechanics, and knew Newton's *principia Mathematica*, which he quoted. The use of Newton's ideas is rendered by the following accounts of Varignon's answer to Rolle: Mancosu quotes the description of Varignon's argument by an academician witness of the debate (Reyneau):

> Puisque la nature des diffierentielles [...] consiste à être infiniment petites et infiniment changeantes jusqu'à zéro, à n'être que *quantitates evanescentes*, *evanescentia divisibilia,* elles seront toujours plus petites que quelque grandeur donnée que ce soit. En effet quelque difference qu'on puisse assigner entre deux grandeurs qui ne diffèrent que d'une différentielle, la variabilité continuelle et indéfinie de cette differentielle infiniment petite, et comme à la veille d'être zéro, permettra toujours d'y en trouver une moindre que la différence proposée. Ce qui à la manière des Anciens prouve que non obstant leur différentielle ces deux grandeurs peuvent être prise pour égales entr'elles [14].

And Blay (1986) quotes the *Registres des Procès-Verbaux des séances de l'Académie royale des Sciences* (t. 19 f. 312 v-313 r)

> Mr. Rolle a pris les différentielles pour des grandeurs fixes ou determinées, et de plus pour des zeros absolus; ce qui luy a fait trouver des contradctions qui se dissipent dès qu'on fait réflexion que le calcul en question ne suppose rien de tel. Au

[14] Since the nature of differentials is to be infinitely small and infinitely changing till zero, since differentials are but *quantitates evanescentes*, *evanescentia divisibilia,*, they will always be smaller than any given magnitude. Indeed, whatever the difference we can ascribe between two magnitudes that differ by only a differential, the continual and indefinite variability of this infinitely small differential, which is as on the brink to become zero, always enables to find a smaller one than the differential suggested. This proves, in the way of the ancients, that notwithstanding their differential, these two magnitudes can be taken as equals. (my translation)

> contraire dans ce calcul la nature des différentielle consiste à n'avoir rien de fixe, et à decroistre insessamment jusqu'à zéro, *Influxu continuo*; ne les considérat même qu'au point (pour ainsi dire) de leur évanouissement ; *evanescentia divisibilia* [15].

In order to answer Rolle's arguments against the foundations of the calculus, Varignon gave a dynamic explanation that drew on Newton's fluxion. He justified operations with infinitesimals with the intuitive idea of continuously decreasing and vanishing quantities, which is the intuition that sustains the concept of limit. However, the realistic ontology about infinitesimals may have hindered for some time the development of the operative notion of going to the limit.

The epidemiological question is: Why did Newtown's theory of evanescent quantities spread in France instead of Leibniz formal theory of infinitely small quantity? This is surprising because the work of Leibniz was the first known in France and because much of it, such as its notations, was taken on by French mathematicians. According to models of cultural evolution, "biased transmission" is what importantly happened in the introduction of the infinitesimal calculus in France. Biased transmission captures the critical role of Malebranche. But what about the somewhat seditious introduction of Newtonian ideas in the infinitesimal calculus? It seems that neither the prestige nor the spread of the Newtonian calculus in France can explain why Newton's ideas would concurrence so successfully the ideas of Leibniz. If the introduction of Newtonian ideas about the calculus cannot be explained in terms of source-based bias transmission, then they may be explained in terms of content-based bias transmission. Sperber and Claidière (2007) give the following example of content-based bias transmission:

[15] Mr. Rolle has taken the differentials as fixed or determined magnitudes and, moreover, for absolute zero; this led him to find contradictions that disappear as soon as one thinks that the challenged calculus does not presupposes this. On the contrary, in this calculus, the nature of the differentials consists in having nothing fixed, but incessantly decreasing till zero, *Influxu continuo*; that shall be considered only when they disappear; *evanescentia divisibilia* (my translation).

> Imagine a comedian telling two new jokes one evening on a television show. Both jokes are much appreciated and adopted by the same number of viewers for future retellings. However joke 2 is harder to remember than joke 1, so that, say, 80% of the people who adopt it forget in less than a month, whereas only 20% forget joke 1 in the same period. Quite plausibly, joke 1 will spread and become a standard joke in the culture, and joke 2 won't. To model such a plausible evolution one should take into account not only frequency of adoption but also frequency of forgetting.

The French mathematicians of the beginning of the 18th century may have been in a situation comparable to the viewers of the television show. They have as input, not two jokes, but two different ideas of the infinitesimal calculus. One of these two ideas is not more difficult to remember, but it is more difficult to think with. Content based biases, say Sperber and Claidière (2007), "are effects of the cognitive mechanisms that construct a mental representation on the basis of informational input." In the previous section, I have argued that Leibnizian infinitesimals are harder to think with than Newtonian fluxion and evanescent quantitites, because the latter still rely and make use of the number sense. The cognitive mechanism from which the bias result is the number sense, and the informational input are theories and application of the infinitesimal calculus. The bias toward Newtonian ideas is partly due to the innate endowment and structure of the mind. With this historical case of the infinitesimal calculus, we find an example of a psychological factor of attraction. The attraction is caused by an ability in naïve mathematics to understand continuous quantities; the cultural representation attracted is the concept of infinitesimals, it is attracted towards notions resembling the concept of limit. This is, I think, the most reasonable thing we can say about a process of "platonistic rediscovery." I have analysed one particular course of events in the history of the numbers, but the number sense has probably had a pervasive influence in the distribution of numerical and arithmetic representations. For instance, De Cruz (2005) hypothesises that the difference in the distributions of positive integers, which have emerged independently in many cultures, and zero, which has evolved only once as a true numerical concept, is due to the relations that these concepts have with the number sense.

7.4 Conclusion: historical analysis and cognitive hypotheses

The epidemiological framework raises the question: **Why do some concepts stabilise so as to enter the corpus of mathematical knowledge?** Here are some possible answers that can be explicated in the epidemiological framework:

- A concept can spread among a population (of mathematicians) only if the *structure of communication* allows it. That is to say the distribution of the mathematical public representations furnishes sufficient input to the minds of mathematicians, who then construct their own representation of the meaning of the public representations. E.g. network of scientists communicating their results such as the network that Malebranche entertained with Leibniz on the one hand, and a group of French mathematicians on the other. This network allowed the constitution of a group of mathematicians -'the infinitesimalists'- that promulgated the calculus in France.
- The efficiency of communication is attained under several conditions, among which we can find:
  - The use of mathematical terms and the development of mathematical ideas rely, at bottom, on *mental mechanisms* through which one can reason with the terms. Cognitive processes can, with such input, build an adequate mental representation. E.g. 1) The public representations for numbers are understood when

associated with mental representations of magnitudes 2) The 'evanescence' metaphor for infinitesimals.

- *Deferential behaviour* is also a key aspect for the stabilisation of a concept. The source needs to be trusted. E.g. Malebranche, after some reticence, came to trust Leibniz on Mathematical topics. L'Hôpital was somewhat a disciple of Malebranche who introduced him, via J. Bernouilli, to the calculus. Fontenelle, using the power and prestige that his position of secretary at the Académie conferred to him acted as the eulogist of the calculus at the Académie Royale des Sciences
- *Contextual interests and background knowledge*. E.g. The success of the calculus can partly be explained by the fact that it increased drastically the predictive power of mechanics.

- The rigor of Mathematics is to be explained with mathematical practices, and more particularly, the use, nature and production of public representations. Here are some aspects of this point:
  - *Autonomy of mathematical notions.* e.g. In order to develop, the Calculus first needed to emancipate its notion of infinitesimals from theological connotations.
  - Public representations are *written* and there is an extensive use of Mathematical *symbols*. This plays a role in decontextualisation, the use of the memory, and allows some important practices that define the rigor of mathematics, such as always going back to the definition.

I provide this list as a contribution to the understanding of the richness of mathematical practices, which are made of social as well as cognitive events. The following graph 7.2 represents a "link" in the causal chains that constitute cultural phenomena. I include some of the elements of the list concerning mathematical cognition and the evolution of the infinitesimal calculus. The point of the graph is that sociology of knowledge and cognitive psychology take turns in explaining what is happening — the causally related events being either mental or taking place in the environment. In some cases, the causes can be both social and mental, as in the role of background knowledge, interests and motivations in the production of new representations. The interaction between the environment and the mind's processes is also often very tight: this is what studies in situated and distributed cognition show. The consequence is that social and cognitive studies of science and mathematics should be tightly integrated.

This chapter asks questions to cognitive psychologists that are relevant to historians of mathematics, and reciprocally, it asks questions to historians of mathematics that are relevant to the studies of the cognitive foundations of mathematics. Using an appropriate theory of cultural evolution — the epidemiology of representation — is what enables asking these interdisciplinary questions, bridging studies about the mind and studies about developing practices. It is also an argument against loose descriptions of the relation between mathematics and the human mind.

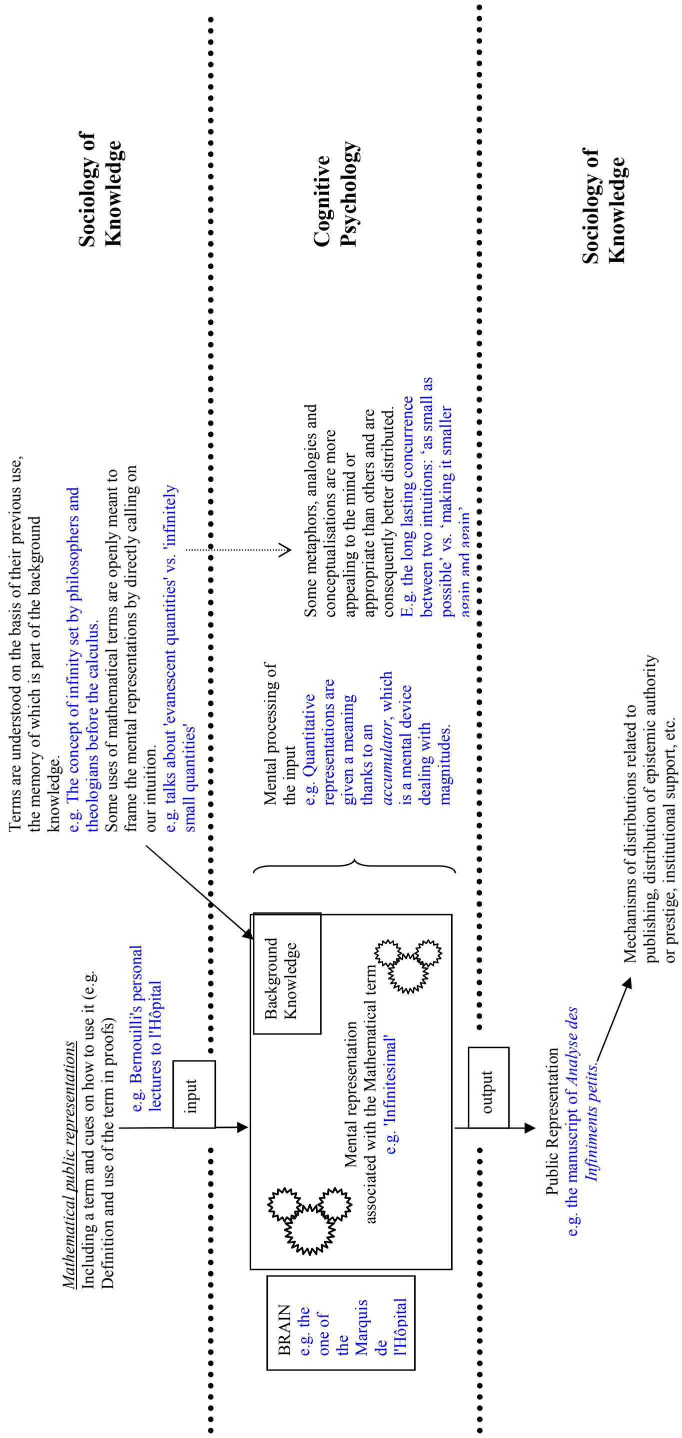

Figure 7.2: mathematicans' minds, the environment, and the cultural causal chains that span through them, to make mathematical knowledge

Part III

The cultural organisation of scientific cognition

Integrating social and cognitive studies of science implies socialising the old image of the scientist working in isolation, but without throwing the scientist's mind with the individualistic philosophy. In the previous chapters, I have insisted on communication, enculturation, assessment by the scientific community, and other social processes that are constitutive of the production of scientific knowledge. It is within such a socialised theory of scientific knowledge that I have specified how properties of the human mind determine the evolution of scientific knowledge. The socialised theory of scientific knowledge I now want to discuss insists on the embedding of scientific practices in material culture and social structures; it says that scientific knowledge is produced by institutions of organised working scientists and artefacts; it asserts that science and technology evolve together and cannot be studied separately. The 'practice turn' in science studies (Pickering, 1992) has emphasised that scientific knowledge comes from scientists at work. Scientists are doing and making things. Their contribution to science in the making cannot be reduced to their choices of believing this or that scientific theory. In particular, sociologists of science have addressed the question of technological evolution. The result is a reappraisal of the role technology plays in the evolution of scientific knowledge, and an application of the theoretical apparatus of science studies (such as the notion of interest) to the study of the evolution of technology (Bijker et al., 1987). The laboratory is the archetypical example of an institution where technology and social organisation play a crucial part in the production of scientific knowledge. A lab is a social entities entity, and thus the product of socio-historical processes. So sociology is necessarily fully implicated in their analysis. One of my purposes will be to point out how psychology can participate in the study of these socio-historical processes. The practice turn has, indeed, demised the interest in the processes of belief formation, which was previously the focus of the Strong Programme. I will argue that understanding scientific practices and the social and environmental embodiment of science require understanding the beliefs scientists have about how to do science — e.g. which tools to use.

In this part, I follow the trend initiated by Giere and Nersessian, and I argue that the production of scientific knowledge is the output of distributed cognitive systems. The first chapter is a critical account of the use of the notion of distributed cognition in science studies. It argues that distributed cognitive systems are adequate descriptions of institutions where the cognitive labour is divided between human and non-human cognitive agents; and it emphasizes the social and the mental aspects of distributed cognitive systems. The second chapter tries to clarify which are those mental and social aspects, and how they can be accounted for. I question the ontological status of distributed cognitive systems. What sort of social entities are they? I argue that they are social institutions endowed with some cognitive functions, and I explain what this definition implies for distributed cognition systems, for their existence and their evolution. The third chapter deals with the evolution of distributed cognitive system as being a key aspect of the history of science. It analyses how distributed cognitive systems can arise out of human mental representations and social behaviour, and eventually focuses on the role of mental representations of trustworthiness. Trust is both what sustain distributed cognitive systems thought time and what cause them to change. The last chapter is a case study in the history of mathematics: it analyses how and why mathematicians came to trust computers for proving theorems.

Chapter 8

Distributed cognitive systems in science

A strategy for integrating social and cognitive studies of science is as follow: first, recognise that scientific cognition extends beyond the borders of the scientists' skull. This is done thanks to advances in the study of situated cognition. Scientific cognition extends beyond the borders' of the scientists' skull, but it extends in a highly structured, culturally framed, environment. So the second step is to describe the highly structured environment within which the scientists work, and the role of this structure in the production of knowledge. This is made possible by the notions of *distributed cognition* and *distributed cognitive system*. This strategy has been advocated most forcefully by Ronald Giere and Nancy Nersessian.

In this chapter, I critically review the work that has been done about the distribution of scientific cognition and distributed cognitive systems in science. In the first section, I introduce the theory of distributed cognition, then I successively analyse the work of Latour, Giere and Nersessian. I conclude on the prospects and limits of the analyses in terms of distributed cognition.

8.1 The idea of distributed cognition

The notion of distributed cognition is best elaborated in Ed Hutchins' seminal book Cognition in the Wild (1995). In this book, Hutchins describes how the task of piloting is performed by a team of sailors in a U.S. navy ship. Piloting consists in navigating near land, especially when coming into port. Hutchins provides a classical description of the culture of the boat, the social organisation and, eventually, the division of labour. He points out that the team works collaboratively and manipulates several artefacts. The important twist, however, is that the task and the labour are described as cognitive, i.e. as concerned with the transformation and production of representations in order to provide important information for the action of directing the ship. The division of labour is then described as division of *cognitive* labour, and each agent is given a cognitive task to perform. The general cognitive goal analysed by Hutchins consists in determining the position of the ship. More specifically, the goal is to represent the position on a chart by the intersection of two lines drawn using bearings from two sightings on opposite sides of the ship. A simplified account of the computations performed is as follow: two sailors are on each side of the ship and must record angular locations, with alidades, of landmarks to the ship gyrocompass; the number resulting from the operations is then communicated, via the telephone to the navigator, who plots the location of the ship on a chart, using a hoey (a kind of ruler). Thus, cognition takes place in the whole process, and along an information flow that goes through several individuals and artefacts. The cognitive task could not be performed by one individual, since one cannot be at the same time on each side of the ship and writing on the chart, and time constraints are important aboard the moving ship. Hutchins shows that artefacts are fully implicated in the cognitive processes: the gyrocompass, the alidade, the telephone, the chart and the hoey are participating in the creation, transformation and transmission of representations of the ship's spatial relationship to known landmarks. Manipulating these representations leads to position fixing. Hutchins' explanation of the cognitive processes implemented in piloting take into account the social structure and culture of the navy ship, which determines the flow of information among individuals, and the artefacts, whose cognitive function can be specified as well as that of human agents. The computation is accomplished "by the propagation of representational state across a series of representational media".

There are several points that Hutchins wants to illustrate with his case study. First and foremost, Hutchins argues that a proper analysis of cognitive processes should not be bounded to what happens within the brain. It is useful, he says, “to adopt a concept of computation that does not require a change of theory to cross the skin”. What he means is not that the brain’s cognitive processes have no specificity worthy of investigation with proper means (e.g. neuroscience); rather, he means that cognitive processes, as such, should be analysed in terms of flows of information and functional relationships between elements that participate in the process. This unconstrained concept of cognition enables the analyst to understand the cognitive function of artefacts, as when the sailor manipulates the alidade and the gyrocompass to issue a number which represents spatial relations. Hutchins complains that traditional, ‘internalist’, cognitive science has failed to see that many of the processes they intended to account for in terms of brain processes, are implemented through the brain and the bodily interaction of the agent with his environment and the processes that happens in the external environment. This failure, explains Hutchins, has led to “overattribution”: “When one commits to the notion that all intelligence is inside the inside/outside boundary [of the skull], one is forced to cram inside everything that is required to produce the observed behaviors”; one “mistakes the properties of complex sociocultural systems for the properties of the individual mind” (p.355). The attribution problem consists in finding out the exact cognitive system that accounts for the observed behaviours. In Hutchins’ case-study, the cognitive system accounting for plotting is made of the navigation team and their tools (in my simplified account: two sailors and their tools and the plotter with his own tools). The physical arrangement and the social-cultural order also determine the operations of the cognitive system. For instance, it is important that the charts be arranged in such a way that they can be readily used when wanted. They are consequently piled on the plotter’s desk on top of one another in order of expected need. Also, the culture of the U.S. Navy plays a role in ascribing different authorities and a domain specific distribution of responsibilities. The navigator, the pilot, the ‘quatermaster’, all have their specific task, for which they are accountable in a somewhat hierarchical social order. Hutchins also describes cultural variations in the cognitive computations that are implemented for the task of navigating. The description in terms of cognitive distributed systems, however, applies across cultural variations. The Micronesian navigation is done through some culturally elaborated representational assumptions, which underlie the use of a representational media for computation, with notions such as the star path, and cognitive tools, such as the sideral compass.

The methodological consequences span large. One first consequence is that cognition should be studied in the wild. This is because human cognition is normally in constant interaction with its social and physical environment. Human cognition makes the most of the environment, and this environment has been culturally shaped. Hutchins warns us that little is known about the relation of cognition "in the captivity of the laboratory to cognition in other kinds of culturally constituted settings." But most of what we know about cognition, he laments, was learned through laboratory experiments. Studying cognition in the wild implies describing the "cognitive task world." Indeed, given a cognitive goals, the cognitive task of the brain is specified by the environment, and especially by the available cognitive tools, which are put to work to attain the goal. For instance, adding large numbers with pen and paper requires different mental operations than adding large numbers with an abacus. Solving the attribution problem, one realises that cognitive processes may involve coordination between internal and external structures, rather than a direct internal confrontation with the task. So a major work of the cognitive ethnographer is to describe the specific cognitive functions of the elements of cognitive systems. For instance, the sub-system made of the plotter, the chart and the hoey, takes as input two numbers and gives as output a dot on the chart. It does that through a specifiable set of actions or processes: manipulating the hoey and the chart in a proper way. When I refer to cognitive processes that do take place in the brain, I will specify and talk about 'mental cognitive processes', since 'cognitive processes' can refer to cognition happening between agents communicating and when manipulating cognitive artefacts.

Some researchers in sciences studies have acknowledged that the framework of distributed cognition can be fruitfully applied in science studies. Latour (1996) praises Cognition in the Wild for the picture it gives of cognition as happening in social and cultural settings rather than just in the head of some individual. He also applauds Hutchins' analysis of the role of artefacts in cognition and shows his keenness for the idea that representational media go across the skin. Latour concludes his review of Cognition in the Wild with regret that Hutchins did not extend his theory of cognition to the production of scientific knowledge, and asserts that history of science would benefit from Hutchins' sophisticated understanding of context and his definition of distribution of cognitive tasks. Similar appraisals come from the cognitive studies and philosophy of science. Nersessian (2005), Giere (Giere, 2002a; Giere and Moffatt, 2003) and Thagard (1993) see in the notion of distributed cognition the means to integrate cognitive and social studies of science. Giere and Nersessian have also actually described laboratories as distributed cognitive systems: the European Centre for Nuclear Research (CERN) (Giere and Moffatt, 2003), the research complex of the Hubble Space Telescope (Giere, 2003), and a research lab in Bio-medical engineering (Kurz-Milcke, Nersessian, and Newstetter, 2004).

Nersessian's cognitive history, Latour's ethnography of laboratory life (that he opposes to "the stuffy atmosphere of epistemology"), and Hutchins' cognitive ethnography have the same project: studying knowledge in the making, and the processes involved as taking place in their natural settings. All three programmes are based on the same assumption that our understanding of science, local knowledge and techniques shall benefit from field work. Field work, they assume, shall provide new data and shall lead to an understanding of actual social and cognitive phenomena that cannot be acquired through the practice of experimental psychology or philosophy.

It is worth reflecting on the differences between Latour's and Nersessian and Giere's interpretation and use of the theory and framework of distributed cognition. For the two last authors, distributed cognition is a notion that provides the methodological means for integrating cognitive and social studies of science. For Latour, distributed cognition makes cognitive anthropology adhere to some principles of his own theorizing.

8.2 LATOUR: DISTRIBUTED COGNITIVE SYSTEMS WITHOUT HUMAN MINDS

Well before Hutchins' book, Latour (1986) has argued for the importance of artefacts and bodily activity in scientific practice. In this article, which is quoted in Hutchins' book, Latour shows the extent to which the scientific practices rely on visual artefacts, and argues that most of the scientist's work consists in the manipulation and production of visual artefacts. Latour shows the relevance of Goody's analysis of literacy and its implications regarding thinking (Goody, 1977) for understanding the scientific revolution as well as current scientific development. The savage mind, he says, is continuously being domesticated with the renewal of means of presentation of evidence. For instance, Latour quotes at length Eisenstein's claim (1980) that Copernic and Tycho Brahe's work in astronomy was rendered possible by the invention of the printed books, which enabled them to access large astronomical data. In other word, Latour is an early advocate and contributor to the analyses of the distributed aspects of scientific cognition.

Because Latour's theorizing includes important elements that resemble some arguments developed by the proponents of distributed cognition, it is important to emphasize an essential difference, which appear clearly in his review of Cognition in the Wild. For Latour (1996), the book shows that one can do without cognitive psychology, and his review is taken as an occasion for him to renew his notorious claims on cognitive psychology: "Nothing, absolutely nothing of what is considered essential to the very existence of psychology is left in the book", he says. According to Latour, one can conclude from Hutchins' book that "cognition has nothing to do with minds nor with individuals" or that "there is no meaning in asking what is in the mind of the plotter". Latour says that "there is not, according to Hutchins, any meaning in the expression 'I think' or 'I represent'." Obviously, Latour pulls Hutchins on his side and his review of Cognition in the Wild is mostly an occasion for him to promote his own view on mind and science. Contra Latour, one does need to take a look at what is happening in the users' head. This is true in the framework of distributed cognition as in any other theoretical framework for the study of culture and society. The importance of internal cognitive processes is apparent in Hutchins' entire book, even if Hutchins himself abstains from specifying the nature of internal, i.e. mental, cognitive processes (although he expresses distaste for symbolic cognition of computational cognitive psychology and affection for pattern recognition in connexionist models). Strangely enough, the very sentence Latour chose to exemplify what he takes to be Hutchins' denial of internal cognitive processes, implies precisely the opposite : "each tool presents the task to the user as a different sort of cognitive problem requiring a different set of cognitive abilities or a different organization of the same set of abilities" (Hutchins, 1995, p.154, quoted in Latour, 1996). Here, the cognitive abilities appealed to are that of the individual user of the tool; they are mental cognitive abilities. And again: "the task performer can transform the task to an easier one by achieving coordination with an internal artefact: the knowledge of this technique"; "these tools permit the people using them to do the tasks that need to be done while doing the kinds of things people are good at: recognizing patterns, modelling simple dynamics of the world, and manipulating objects in the environment" (Hutchins, 1995, p.144 and 155, quoted in Latour, 1996). Humans are important elements of distributed cognitive systems, so it is not surprising that the cognitive processes they implement are themselves important. Moreover, what is pregnant in Hutchins's approach to cognitive tools and distributed cognitive systems is that they are designed so that humans can achieve what they aim at, while doing the things they are good at: cognitive tools and systems are built so as to make the most of the limitation of human cognitive abilities. "Humans create their cognitive powers by creating the environments in which they exercise those powers" (p.169), Hutchins says; and the environments they create are precisely physical artefacts as cognitive tools, and social organisations as cognitive distributed systems.

Of course, Latour is not blind to the fact that Hutchins has a different agenda, concerning cognition, than his own. Latour, indeed, blames Hutchins for willing to improve and advance cognitive psychology. Latour's criticism of cognitive science is fed on the idea that:

Psychology is not there to describe events but precisely to cram cognition inside an individual mind [...] To believe that a better cognitive science will simply take over, is to miss the anthropology of the moderns and to underestimate the history that made the myth of the internal state so essential to our Occidental life.

Latour's provocation and his claim about the nature of psychology laid aside, there is both a fair ground to Latour's denial of the explanatory power of cognitive psychology, and an important mistake underlying his interpretation of Hutchins' book as a permission to dispense with psychology. The fair ground is constituted by the assertion that the scientist's mind is in nothing different from the mind of lay people, and, more radically, that scientific thinking has no specific feature that could make it scientific. Good reasoning, for instance, can be found in many non-scientific activities. Such assertions are opposed to philosophical attempts at characterising rationality as what would make science special. In brief, Latour argues (1) that what distinguish science from other activities is not a special properties of scientists' mind, and conclude (2) that psychology is useless. The premise (1) is correct. Latour's argument is based on the history of science: the scientific revolution did not happen because of the birth of some new kind of humans, endowed with new cognitive abilities for thinking scientifically (1986). But the conclusion (2) does not follow, and is in fact erroneous. If psychology is useless for providing a criterion that distinguish science from non-science, this does not make it useless for the description and understanding of the scientific activity. The main goal of science studies is to understand naturalistically how science is produced rather than pursuing the old agenda of finding a demarcation criterion.

The scientist thinks whether or not this thinking is what makes him a scientist, and his thinking is an integral part of his scientific activity. The fact that there is nothing specifically scientific in a scientist's cognitive processes does not mean scientists do not have minds to make science with.

The 'historical argument' presented above asserted that the specificity of scientific thinking cannot be found in the mind, the 'ethnographic argument' continues the proof by showing with actual cases the importance of what happens outside of the mind for the production of scientific knowledge (Latour, 1986). Latour sees in Hutchins' book a confirmation of his argument against psychology of science because the notion of distributed cognition seems to displace the locus where representations are transformed and produced from mind to environment. Here again, the premise is well argued and illustrated: there are indeed events that take place outside of scientists' brain and that are constituents of scientific cognition. But the conclusion against psychology of science does not follow. The step from the premise to the erroneous conclusion relies on a misunderstanding of the role of the mind in distributed cognitive tasks. The error consists in thinking that if cognition is outside of the brain, then it is not inside. It consists in thinking that distributed cognition implies light mental cognition that one can dispense with. Latour describes this human cognitive agent as:

A very lightly equipped human agent [. . .] like the actor of ethnomethodology [. . .]. Instead of cramming endless numbers of modular boxes and special purposes rules in the head, Hutchins takes everything out and "renders to Caesar what pertains to Caesar"

There is more to distributed cognition than merely transferring cognition outside the brain Hutchins argues on the contrary that cognitive ethnography contributes to cognitive *psychology* by specifying what tasks are imparted to the individual in a normal environment (outside the lab). And if Hutchins asserts that what was thought to be inside the head is in fact outside it, he also repeats that the consequence is that something else than what classical cognitive science had hypothesised is actually going on inside the head. The individual happens to strongly rely on, and inter-acts with, the environment for the achievement of his own cognitive goals. The consequent hypothesis should then be that the individual is certainly endowed with the ability to interact in a complex way with his environment, which contradicts the commonly derived conclusion according to which, since the mind relies on the environment, 'there is nothing left in the mind'. As Sperber says, "the richer the interaction of an organism with its environment and with others, the richer must be its cognitive capacities" (2006b). In particular, human agents considered as elements of cognitive systems often must use some tools, which is known to require some evolved cognitive capacities, and they must manage the relations they have with other humans in the distributed cognitive systems, which requires some abilities to manage social relations. Hutchins indeed notices that "wherever computations are distributed across social organization, computational dependencies are also social dependencies. Performance is embedded in real human relationship" (1995, p. 224). With distributed cognition, there is no transfer of cognition from the inside to the outside, there is more cognition.

Moreover, if science largely takes place in the scientists' environment, then an analysis of the development of science should include an analysis of the processes out of which the environment is built up. Studying science shall include the study the construction of the various niches (social, technological, cultural) in which scientific activity takes place. These niches are constituted of the scientists' social and material surroundings. These determine what individuals a scientist can interact with, how and about what, and the cognitive artefacts that are placed at their disposal. In particular, distributed cognitive systems are such niches. I will study the historical processes that lead to the emergence and development of distributed cognitive systems. It will appear that an understanding of these processes benefits from research into the mental representations of the people involved. So cognitive psychology will prove to be useful not only for understanding how distributed cognitive systems work, but also for understanding how the systems emerge and evolve.

8.3 Giere: Distributed cognition is where the cognitive and the social merge

Nersessian and Giere's use of the notion of distributed cognition is much more faithful to Hutchins' idea. Hutchins' prior question was about the contribution that anthropology could make to cognitive science. His answer took the form of a cognitive analysis of socio-cultural phenomena, which leads to rethinking the role of the mind (the biological human cognitive apparatus) in cognition. When thinking about the relations between culture and cognition, Hutchins found that cognition occurs also in socio-cultural phenomena, consequently leading to a criticism of individualistic cognitive science. Giere and Nersessian took the idea of distributed cognition as a means to step out of individualistic cognitive studies of science (see also Thagard, 1993). They both had started to study scientific cognition by focusing on the cognitive processes implemented in the mind of individual scientists. In particular, they had investigated scientific theorising as the mental construction of mental models (Giere, 1988; Nersessian, 2002a, 1999), and Nersessian (1984) investigated the thoughts of particular historical scientists, and the thought processes leading to conceptual change (1992b; 1999; 2002b). Their ensuing question was the role of social and cultural phenomena in scientific cognition. Nersessian still now laments that "cognitive accounts, while paying deference to the importance of the cultural dimensions of practice, have, with few exceptions, not made cultural factors an integral part of the analysis" (2006, p. 125).

Giere first incorporated social factors in scientific knowledge production as non-epistemic interests that determine theory choices of scientists independently of consideration of evidence (Giere, 1988). A move criticised by sociologists of science for its distinction between purely epistemic factors and social factors (Pickering, 1991), and on which he comes back with his discovery of distributed cognition (Giere and Moffatt, 2003). Nersessian, for her part, advocated a cognitive history of science as making justice to the temporal and contextual determinations of scientific thinking (Nersessian, 1995). But they both found in the notion of distributed cognition a means to further integrate social and cognitive studies of science.

Giere and Moffatt present distributed cognition as the place "where the cognitive and the social merge" (Giere and Moffatt, 2003). They themselves oppose this view to Giere's previous understanding of the social factors involved in producing scientific knowledge, where "any social story was viewed as distinct and, indeed, supplementary to, the cognitive story" – a view they claim was the standard view among cognitive scientists of science. But if one observes how cognition is distributed in scientific practice, then "the cognitive and the social merge precisely because we cannot say how the scientists work together to complete their cognitive task without describing their social interactions" (Giere and Moffatt, 2003). For instance, Giere (2002a) says about the Cyclotron Facility Indiana University that it consists of a distributed cognitive system that includes the accelerator, detectors, computers, and all the people working on the experiment. About the social relations, he then adds:

> It is not irrelevant to the operation of this cognitive system that the people monitoring the data acquisition are most likely the PhD faculty members who participated in designing the experiment. Those tending the detectors may have PhD's, but may well be full-time laboratory employees rather than faculty. Those keeping the accelerator in tune are probably technicians without PhD's. One cannot adequately understand the operation of the whole cognitive system without understanding these differences in roles and status among the human participants.

Giere and Moffatt (2003) also provides a straightforward application of the notion of distributed cognition to the description of the Hubble telescope. The framework is used to trace the flow of information from the observed cluster of galaxies Abell 1689 to the theoretical claims output of Space Telescope Institute in Baltimore. Also, Giere and Moffatt (2003) and Giere (2002b) wittingly reinterpret Latour's (1986; 1999a, chap. 2) and Knorr-Cetina's (1999) analyses into the distributed cognition framework, and thus strip off Latour and Knorr-Cetina's analyses of their unnecessary metaphysical add-ons: Latour's denial of the distinction between the human and non-human agents, and Knorr-Cetina's attribution of agency and knowledge to social entities. Giere (2004) criticises Clark, Hutchins, and Knorr-Cetina's attribution of agency to distributed cognitive systems as "extensions [that] do not provide theoretical advantages for the study of science. On the contrary, they introduce a host of theoretical problems that confuse rather than enlighten. We are theoretically better off rejecting these supposed innovations" (p. 768). Giere's motives for the rejection of these 'supposed innovations' are based on our lay notion of human agency, which comprise notions that can be found in cognitive psychology – such as intentions, ability to plan, having memories, etc. – but also more 'moral' notions – such as responsibility of human agents. For a naturalistic study of science, however, the specificity of the human mind as described by psychology provide a sufficient reason against metaphors that further blur the notions of mind or consciousness, or that drown human cognitive specificity into the undistinguished term of 'actant'. Further appeal to moral notions, such as problems of responsibility, are unnecessary and makes the argument tumble under Latour's criticism against psychology of science, in which he perceives a means to reintroduce Cartesian dualism.

Recognizing the specificity of human cognition is essential for the integration of cognitive and social studies of science. This is because the fruitful prospects of such integration mostly reside in the knowledge about the human mind that comes from cognitive science. In particular, I shall point out that there are, in situations where cognition is distributed, specific human abilities that enable the distribution of cognition and the maintenance of distributed cognitive systems: human agents can have representations of some artefact or person as having a cognitive function; such representations, if well distributed, can cause the creation and maintenance of distributed cognitive system. It is, indeed, on the basis of such representations that scientists decide to use an artefact or a person for the achievement of a cognitive goal, and thus organise the distribution of cognitive labour.

Giere's criticisms, however, do not only bear on the metaphorical extensions of the notions of mind and agent. His translation of Latour's articles (1986; 1999a, chap. 2) in terms of distributed cognition is also taken as an opportunity to argue against relativism in social studies of science. Giere conclude his analysis of Latour (1986) with:

Results do not come to be regarded as veridical because they are widely accepted; they come to be widely accepted because, in the context of an appropriate distributed cognitive system, their apparent veracity can be made evident to anyone with the capacity to understand the workings of the system.

The first striking point in Giere's argument is that his view, presented in the second part of the quote, does not necessarily contradict Latour's view, as presented in the first part of the quote. Events in the history of science can happen as follow: In the context of a distributed cognitive system, a result is produced and is interpreted as making evidently true some statement about the world. In this context, scientists come to 'see' the apparent veracity of the statement, and this causes them to be convinced. The scientific community then accepts the stated description of the produced result. This wide acceptance provides the statement with the status of being a scientific truth. In fact, that it is (one way to present) Latour's point: scientists operate a series of 'translation' which bring about a power to convince; translations include simplifications that make the result evident. As for the relativists targeted by Giere's attack, they have no claim about what make a statement true: relativists think it is the task of scientists to decide what make their assertion true in their own field; it is *not* the task of scientists of science. Relativists question how some claims come to be accepted as a true scientific claims and they try to answer without postulating on the truth of the claim. This leads me to my second point about Giere's argument: Giere specifies that the distributed cognitive system that forms the context of persuasion has to be "appropriate". This makes the perfect question for the relativist: who can decide when a distributed cognitive system is appropriate, and how? In the next chapter, I describe a case in the history of mathematics where the question 'What is an appropriate distributed cognitive system?' is asked by the scientists themselves, and I analyse the factors that determine their decisions. As it was asked in the historical debate, the question is: 'can computer assisted proofs generate genuine theorem?' Which translates as 'are distributed cognitive systems that include computers as elements appropriate systems for the production of mathematics?'

Giere's use of the framework of distributed cognition does not lead him to raise questions in the sociology of scientific knowledge. The notion of distributed cognition allows him to encompass the social within the cognitive. In his accounts, the social aspects of science are described only to specify the flow of information and the division of cognitive labour. The social organisation of the CERN, for instance, seems to answer one question: since one cannot do nuclear experiments on one's own, how can we organise the (social) repartition of tasks so as to perform it in the best conditions? This kind of question is a step forward with regard to Mertonian sociology of science, where social aspects are thought as only hindering or facilitating an otherwise purely rational development of science. With distributed cognition, Giere upgrades social phenomena from *enabling conditions* of science to actual *components* of scientific practice. But social-historical context and cultural background still do not appear as part of what determines scientific knowledge. The social phenomena taken into account are only those that fulfil some rational social plan. As for any rational reconstruction, the danger is that what is taken as rational is anachronistic or ethnocentric. An alternative to rational reconstruction questions what is taken as rational by the historical actors, and see if their representations of what is rational has indeed determined the form of the distributed cognitive system.

Lastly, an important contribution from Giere to the study of distributed cognition in science is to show that the history of science may be fruitfully described as a history of distributed cognitive systems for scientific knowledge production. I will devote one section to this point (10.1, p. 299), so I do not develop it now.

8.4 NERSESSIAN: EVOLVING DISTRIBUTED COGNITIVE SYSTEM

Nancy Nersessian and her colleagues - Elke Kurz-Milcke, Wendy Newstetter and Jim Davies - have used the notion of distributed cognition to account for knowledge production in a bio-medical engineering laboratory, where they have been conducting field-work since 2003. The team of cognitive scientists thus confronted collaborative work in the actual site of scientific activity, which corresponds to their willingness to deal with both social and cognitive aspects of scientific knowledge production. Nersessian (2005) introduces the teams' analyses by a review of the developments in cognitive science that show how deeply cognitive and social phenomena are interrelated. These developments, that she coins "environmental approaches," show that human cognition includes social aspects that must be accounted for in cognitive studies of science. These approaches also provide a view of cognition which is far away from the view attacked by social studies, and answer the qualms of those who saw cognitive studies as renewing the erroneous theory of Cartesian dualism. Hutchins' analysis of distributed cognition figures, of course, as an 'environmentalist approach', and puts together some important hindsight of the approaches, such as the cognitive reliance on artefacts and the implementation of cognition in social systems. Nersessian disapproves of the artificial divide between social and cognitive studies of science, and raises the "integration problem": How can one integrate approaches that have been artificially divided, that have developed separate accounts of scientific development, and that sometimes radically oppose each other? The environmental approaches to cognition, says Nersessian, "offer significant groundwork for thinking about the integration problem." The environmental approaches emphasise the pervasive reliance of cognition on the environment: the 'embodiement' of cognition, its 'social embedeness' and its 'situatedness.'

The essential reliance on the environment in scientific cognition is shown in Kurz-Milke, Nersessian and Newstetter's (2004) article, where they consider 'simulative model-based reasoning' as it occurs not only in the head of the scientists, through mental models and thought experiment, but also in the acts of designing and engineering physical devices, or "benchtop modelling" that emulate some natural occurring conditions. This leads the authors to talk about "models-in-action", in order to emphasise that models are also in the action of using the lab's devices as cognitive artefacts. For instance, the "flow loop" is a device that represents blood vessel walls and that is manipulated in the generation of knowledge. The dense and dynamic relations holding between cognitive modelling processes – which include building mental models and physical devices, the understanding of these devices' functions and references, and the actions of re-engineering – is coined "fabric of interlocking models" by the authors. Using a "mixed-method approach" that combines cognitive ethnography and cognitive history, Nersessian and her colleagues, conclude that research laboratories are best qualified as evolving distributed cognitive systems. Cognition in innovative, creative settings, where artefacts and understandings are undergoing changes over time, they say, call for a diachronic understanding of laboratory distributed cognitive systems.

Nersessian advocacy of a cognitive history of science (1995), for the integration of the study of the historical conditions with the study of the cognitive processes involved in scientific discovery, is already a major step in the direction of integration. With the notion of *evolving distributed cognitive system*, she goes one step further by considering that social aspects intervene not only in the historical conditions of scientific discovery, but also in the modes of production of scientific knowledge. The emphasis on the *evolution* of cognitive distributed systems is a key point for applying the framework of distributed cognition to the understanding of scientific development. In their accounts, Nersessian and her colleagues describe the distributed cognitive system of the laboratory they study as being in constant evolution: artefacts, such as the flow loop, are continuously transformed; the 'problem space' is constantly changing as new questions and challenges arise; and there is an important 'turn-over' since the lab has members who stay for limited periods only (such as Ph.D students). An important point of their ethnography therefore consists in explaining this constant evolution and its consequences. For instance, they describe how the flow loop is being redesigned so as to incorporate further constraints of the phenomena it is intended to emulate, or for more technical reasons, such as facilitating experiments and increasing the success rate of experiments. Redesigning is part of the 'mission' of the lab, and is often done in response to problems encountered. Newcomers to the lab have to familiarise with its material culture. Nersessian and her colleagues argue that this familiarisation require hands-on practice with tutoring. More surprisingly, they argue that familiarising with a device implies appropriating some of its history: how and why it has been redesigned.

Although Nersessian and her colleagues give a good sense of how and why the distributed cognitive system evolves, they do not hypothesise on the general cognitive principles that make the evolution possible. As a research lab, the distributed cognitive system has to evolve. In particular, it constitutes a "problem space", where answers imply evolution. The tasks of the researchers includes redesigning the artefacts as elements of the distributed cognitive system. But what is it that generates the changes? What are the mental processes, if any, that ground the actions of researchers changing the distributed cognitive system of which they are a part? My hypothesis is that human elements of distributed cognitive systems have representations about the other elements of the systems with which they interact. According to these representations, they will trust, or not, the elements of the system for some given tasks. So according to the problem they want to solve, human elements will decide to change or re-design, or not, the elements with which they intend to solve the problem. In the next chapter, I will describe the introduction of computers in mathematical practice as an evolution of cognitive distributed system; and I will also search for an explanation of this evolution. I shall essentially rely on the idea mathematicians have about computers and the tasks they can perform.

In some of their writing, however, Nersessian and her colleagues, seem to envisage the relations between artefacts, as cognitive tools, and their human users in a different way. They want indeed to recast "some traditional cognitive science interpretive notions by which we are attempting to break down the internal-external distinction – a major impediment to integrating cognitive and socio-cultural dimensions of scientific and engineering practices" (Nersessian, 2005, p. 41). This assertion could be construed as supporting Latour's symmetry principle, which prescribes avoiding the distinction between human and non-human actors. To avoid such construal, I shall interpret Nersessian's quotation in a direction more favourable to the cognitive psychology of science, to which Nersessian herself contributed.

Why would the internal/external distinction impede the integration of cognitive and social studies of science?

One thing that integrating cognitive and socio-cultural dimensions of scientific and engineering practices could aim at is explaining the social and cultural nature of scientific knowledge production as emerging from the actions of well defined entities: scientists – who are endowed with complex cognitive devices, their minds/brain, and who interact among themselves and with their environment. According to this research program, science is a historical and socio-cultural phenomenon and, as all socio-cultural phenomena, it grounded in, and partially made of, psychological phenomena. The programme aims at clarifying what are these psychological phenomena, and how and why cultural phenomena arise out of them. In this perspective, explaining evolving distributed cognitive systems implies understanding them as organisations that emerge from individuals' behaviour and interactions. Thus, the internal/external distinction, insofar as it refers to the mind/brain as a distinct entity that differs from people's environment, is essential to the integrations of cognitive and cultural studies of science. It is essential because integrating cognitive studies to social studies of science implies that the former do have something to contribute to the latter and/or reciprocally.

The application of environmental approaches for the integration of cognitive and social studies of science does not jeopardise the internal/external distinction either. The point that the environmental approaches strongly make is that the activity of the mind should not be thought of in isolation. In order to make sense of what the mind does, one needs to look at the environment. In order to understand what is internal, one needs to look outside. This is because the mind relies on, and exploits, properties of the environment. Environmentalist approaches criticise an individualistic psychology for which the environment is just a source of input to the mind, and show that it has a greater role in cognition. They develop alternative accounts of cognition that take into account the action of the agent on his environment. Yet, environmentalist approaches still rely on the internal/external distinction. They show that the integration of psychological and social studies are essential, because an understanding of the functioning of the mind must include an understanding of its interaction with its environment, its social and cultural environment in particular. So the internal/external distinction is an impediment to integration only to the extent that it is thought as delimiting two independent realms: the psychological and the environmental. So breaking down the internal-external distinction should not mean rejecting the distinction, it means that understanding of cognition requires understanding many phenomena that cut across the internal-external distinction.

But Latourian ideas still seem to sneak in Nersessian and her collegues' ideas about the interactions between human agents and artefacts. They take position against the idea that interactions between artefacts and humans are lead by human agents' representations about the artefacts. They aim "to construct an account of the lived relation that develops between the researchers and specific artifacts, rather than an account of the developing knowledge about these artifacts per se. By focusing on the lived relations we mean to emphasize the activity of the artefacts in a relational account of distributed cognitive systems." Eventually, Nersessian and her colleagues "characterize the relationships between the various technological artifacts in the cognitive system and the researchers as cognitive partnerships." The metaphor is suggestive of the role of artefacts in distributed cognitive systems: not only do artefacts produce output, they also require specific input from their 'human partner', and their physical properties places constraints on what human partners can and shall do. Yet, the metaphor conceals important differences between the cognitive properties of human agents and the cognitive properties of artefacts. Partnership implies trust between partners. This means that each partner has acquired a representation of the other partner as being a reliable partner. They think themselves as contractually or morally bounded to respect their duty of partner. Such thoughts can also be entertained by the scientists interacting with a device: the scientist believes that such and such device will perform a function he wants to be performed. But cognitive artefacts do not have such beliefs. No cognitive device that I know of is able to evaluate others for partnership. Consequently, the 'lived relations' only metaphorically develops between the researchers and specific artefacts. What really happens is that scientists devise ways in which artefacts can be related to the accomplishment of their goals, and specify their goals as a function of their understanding of the constraints put by these artefacts (e.g. their material properties). Most of the relation between artefacts and researchers is to be accounted for in terms of the knowledge researchers have about the artefacts. As Nersessian et al. say, the knowledge is not about the artefacts per se. But the knowledge is about the artefact as a cognitive tool, i.e. as related to the cognitive goals of the researchers.

Nersessian et al.'s metaphors can, in a Latourian climate, be misguiding: they tend to hinder the study of psychological phenomena that generate distributed cognitive systems. As a matter of fact, cognition, as produced outside the heads of the scientists exists and makes sense only through the goals, beliefs and activity of the scientists. More generally, it is only because artefacts or people are ascribed cognitive functions that they can be described as cognitive elements of distributed cognitive systems and that their activity can be described as cognitive. This is apparent when we contrast the two following situations: When a child plays with an abacus because he likes to move the balls, the movements of the balls should not be accounted for by the mathematical operations they represent. Maybe the child uses the abacus as a kind of rattle. But when an abacus is manipulated by a Chinese seller, then the movements of the balls are explainable as the physical implementations of semantic processes – a ball on the right represents the operation of adding one or five (depending on the ball). The movements of the balls are cognitive processes. This point is recognised by Nersessian when she says that "not all parts of the cognitive systems are equal. Only the researchers have agency and intentions, which enable the cognitive activities of specific artifacts". Here, Nersessian echoes Giere's (2004) worries about the locus of agency. I would insist that what is important is the specificity of human agents' cognitive abilities; I will argue that one crucial cognitive ability is the one of trusting other things or people for specific tasks.

8.5 PROSPECTS AND LIMITS OF DISTRIBUTED COGNITION ANALYSES

The above authors, Latour, Giere and Nersessian, convincingly show that the notion of distributed cognition can be made useful for the analysis of ethnographic and historical data in science studies. Because distributed cognition is both about cognition and about culture (social organisation and material culture in particular), the concept can function as a tool for the integration of social and cognitive studies. In particular, the notion of distributed cognition allows one to describe the detailed interactions of human agents with their material and social environment, and to make sense of these interactions as producing meaningful output.

In the classical view of cognitive anthropology, cognitive psychology is dealing with the cognitive processes with which information is transmitted and transformed, while cognitive anthropology is concerned with the content of the information processed by the minds of the natives. This provides a particular understanding of the internal/external distinction: the internal is the place where information is processed and the external is the place where information is kept and accumulated across generations. A similar view is tempting in science studies, with scientists thinking and producing new representations, while culture furnishes the knowledge acquired by previous generations as input to present thinking scientists. This view is often rendered by quoting Newtown "If I have seen farther, it is by standing on the shoulders of giants". The seeing farther is a genuine cognitive process exercised on the cultural content provided by past giants. In this view, the historical and cultural dimension of science is rendered by a classical history of ideas, together, maybe, with a description of scientific enculturation or education. The historian of scientific ideas looks for which shoulders scientists stand on. The analysis of distributed cognition points the insufficiency of such accounts. The social and cultural environment actively participate to the processes of transformation of representations. With this observation, Hutchins, as well as Sperber (1996a), depart from the classical views in cognitive anthropology and the history of ideas. This is also, I think, against such distinction between the internal as processing knowledge and the external as providing the input that Nersessian argues. Information is not processed within the brain only, but also in the environment and with the environment.

The fact that the processes of production of scientific knowledge are not 'crammed' in the head of the scientists is, to sociologists of scientific knowledge, far from new. Latour is right to see in the environmentalist developments of cognitive science an eventual rejoinder to many of the findings of the sociology of scientific knowledge. And Latour indeed has provided an early contribution to this type of analysis, which, with the 'practice turn' has become classical in social studies of science. So what is new with the distributed cognitive framework? Is it just that some cognitive scientists of science came to discover on their own what was known for long on the other side of the social/cognitive divide? There is indeed a risk that works in the distributed cognition framework say the same thing as, and nothing more than, ethnomethodologists or other practice-oriented studies of science. The difference would only rely on the use of different technical vocabulary and in different historical affiliations, but the content of the descriptions would eventually be similar. The only change is then that practice oriented studies can be now be pursued under the label of cognitive studies

As argued by Giere and Nersessian, the important change relies in the prospects of a real integration of cognitive and social studies of science. But integration is not completed with the mere inclusion of some social studies of science within the set of cognitive sciences (as when social phenomena are shown to be cognitive). Integrating social and cognitive studies of science implies showing the relevance of psychological and cognitive studies of science for social studies of science, and vice versa. It implies working at the compatibility of two arbitrarily distinct fields, and establishing theoretical relations. The works of Giere and Nersessian go in that direction along different trends. In particular, they complement their work on mental models, pertaining to cognitive psychology, by adding to their pictures the role of physically implemented models. They have then sketched the interrelations that exist between mental and material models, rendering the richness and complexity of the relations by appealing to the notion of "interlocking models" (Kurz-Milcke et al., 2004). Giere suggests that, while most abstract models in science are too complex to be stored as mental models, scientists have simple mental models with which they produce "external representations in order to reconstruct aspects of the abstract model relevant to the problem at hand" Giere (2002c, Section6). Craig, Nersessian, and Catrambone (2002) also provide an example of an environmental approach in cognitive studies of science where perceptual context is presented as a problem solving tool in analogical reasoning: a material source of analogy includes perceptual affordances for simulation in the target context. Here again, we have an analysis of how mental processes – drawing analogy, simulating – rely on and interact with the environment – which provide the source for analogical thinking. The environment, of course, is also acted upon, as when diagrams are drawn. In such an integrative perspective, distributed cognition is not only the place where the cognitive and the social merge, it is also a place where theories in cognitive psychology and sociology can meet, confront and enrich each other.

In brief, environmental approaches in cognitive science remain, to a large extent, theories in cognitive *psychology*. These theories draw on the environment in their explanations of cognition because the mind delegates to it much of information processing. Also, the environment is acted upon for cognition: adding large number is done through writing down the numbers; cognitive tools are constructed; social organisations, such as research labs, are instituted. More generally, people construct their cultural environment, which in turns participate to, and thus determine, cognition. There are therefore two questions that arise from environmental approaches of cognition: the first one, which is well investigated, consists in analysing how people use the environment for cognition; the second, which is little investigated, consist in analysing the processes out of which cognitively useful environment is constructed. It is on this second, less investigated question, that I want to dwell on in the rest of this chapter. More precisely, I will question how distributed cognitive systems emerge, are maintained and changed.

The notion of distributed cognition makes possible the *descriptive analysis* of the social division of cognitive labour, but it furnishes no cues about the processes through which distributed cognition emerges. In this respect, it is a limited conceptual tool for the understanding of the social and cognitive aspects of the evolution of science. This point is, of course, not a criticism of the concept, but it points towards the necessity to implicate more sociological theory and more psychological theory for the integrated study of science. For instance, distributed cognition enables describing only one aspects of the situatedness of human agents in society, viz. their cognitive role within some larger cognitive system. But the behaviour of scientists can rarely be explained with the sole description of their cognitive function. In order to grasp the richness of the human cognitive agents, who create, sustain, change and abolish cognitive distributed systems, more psychology and more sociology is needed. This point is apparent when, for instance Kurz-Milcke, Nersessian, and Newstetter (2004) describe how young scientists entering a lab – i.e. starting to take part in a distributed cognitive system - strive to fulfil their jobs requirements. The authors show that it is not sufficient, for these researchers, to grasp their 'job description'. They also have to understand artefacts as cognitive tools, and thus grasp their potential to serve some cognitive functions. They have to understand the social organisation of the lab and its general goal so as to do research as members of a research team. Moreover, these researchers take part in other socio-cultural entities in addition to the distributed cognitive system of the lab. They therefore bring in the lab a set of skills and interests that influence their behaviour as scientists, and thus the knowledge production of the lab. Nersessian (2006) provides an example of such an influence when she recounts how a Ph.D. student, after a visit to some other research institution, brought in some know-how in his home lab.

Nersessian and her colleagues give a sense of the constant evolution of the distributed cognitive systems they have studied, but they do not clearly hypothesise some principles that account for evolution. The hypothesis I defend is that organisational changes in distributed cognitive systems consist in re-ascribing cognitive functions through the trusting behaviour of scientists. Scientists can trust other scientists, or artefacts, or theoretical notions for solving certain cognitive tasks. Trust is a behaviour that is motivated by a complex representation of the thing to be trusted as being trustworthy for solving or helping to solve a given class of task. I argue that changes in the content and distribution among scientists of representations of trustworthiness are a major source of change in distributed cognitive systems.

Emphasising the role of scientists' mental representations lead me to adopt what Latour would reprovingly call an asymmetric position with regard to the distinction subject-object: the scientists do have trusting behaviour towards other people and things, that is best accounted with the *beliefs* they have *about* these people and things. I am also committed to the internal/external distinction regarding cognition, where internal refers to the mind/brain and the body, and external refers to the environment. What distributed cognitive analyses show with regard to the internal/external distinction, I remind, is that some processes that were thought to happen only within the mind, are in fact produced by larger systems of humans interacting with their environment. So the distinction, which belongs to common sense, is not jeopardised by the analyses of distributed cognition: even though cognition cross the boundary of the skin, this boundary is still highly relevant for understanding many phenomena. This is because the cognitive apparatus of the human of mankind is very specific and complex, in such a way that its study deserves a discipline of its own, psychology. In particular, scientists' representations of trustworthiness are internal mental states, and possessing such complex representations as well as processing them (for example using them in planning) seem proper to the human mind. More generally, the distributed cognition framework deserves a renewed focus on human agents for at least the following reasons:

- Humans are components of distributed cognitive systems. The understanding of their cognitive processes is part of the understanding of the processes of distributed cognitive systems.

- Humans provide functions to the artefacts they use, they can change their physical properties or the functions ascribed to them.

- Human behaviour is determined by the social context, which is not reduced to the distributed cognitive system within which the scientific behaviour takes place. Scientists' life outside of their laboratory may be relevant to their behaviour at work and, more generally, their performing of a task within some cognitive distributed system is not fully determined with the specification of this task.

In this section, I have specified my own understanding of the utility and limits of the framework of distributed cognition for the understanding of scientific developments. I have explained my methodological standpoint in relation to the work of Latour, Giere and Nersessian, who had already addressed the question. The conclusion I draw is that the description of how scientific cognition is distributed is an important first step on the basis of which historical developments of science can be explained.

But the explanation of the developments must go forward by providing further attention to human agents as they frame distributed cognitive systems. Accounts of their 'framing behaviour' need take into consideration scientists' general cognitive abilities and their general social contexts. And since social organisations are rarely the product of the design of one individual, but historically develop, in context, through numerous social interactions, the social processes that sustain these historical developments should also figure in the account. Who or what is fixing the cognitive goals or tasks of distributed cognitive systems? According to which historical processes are distributed cognitive processes framed? Who has the power to distribute the cognitive labour? These questions all relate to the processes through which distributed cognitive systems are designed.

Chapter 9

The social organisation of cognition

In the previous chapter, I insisted that cognitive psychology — i.e. the study of the cognitive processes that happen within the mind/brain— has an important role in explaining distributed cognition. Sperber (2006a) warns against the temptation of doing without psychology that stems from an erroneous understanding of distributed cognition.

> Among the few anthropologists who do pay attention to what is happening in cognitive science, there is an often a great readiness to favor what might be described as 'low' or 'light cognition' (on the model of 'low cholesterol' or 'light beer'). Anthropologists are attracted to 'situated', 'embodied', or 'distributed' approaches of cognition. So am I. Indeed, students of culture cannot but be particularly interested by insights into what connects the individual to the environment and to others. Anthropologists, however, seem to like these relatively novel approaches (just as they liked, a few years ago, connectionism) not just for these good reasons, but also because they assume that more situation, body, and distribution means less or lighter mental stuff. Here I disagree. The richer the interactions of an organism with its environment and with others, the richer must be its cognitive capacities

In this chapter, I provide a specific illustration of the need to take into account mental representations by showing their role in the creation and maintenance of distributed cognitive systems. I provide an analysis of the nature of distributed cognitive systems with the framework of the epidemiology of representations. In the first section, I specify the reason why it is interesting to talk about 'systems': systems have emergent properties of their own. I then explain the role of 'regulatory representations' in the organisation of cognition in distributed cognitive systems, and what it means for a distributed cognitive system to have a cognitive function. Eventually, in the last section, I provide a naturalistic characterisation of distributed cognitive systems.

9.1 Emergent properties of distributed cognitive systems

Using the notion of distributed cognitive system implies several points.

- *First*, it is asserted that cognition is distributed; cognition is not crammed within a solitary and disembodied mind. This point is most developed in the environmental approaches in cognitive science, and it is the point that is promptly taken on by Latour.

- *Second*, there is the assertion that the distribution of cognition constitutes systems. The contention of the distributed cognition framework is not just that cognition is leaking out of people's minds (using Clarke's expression, 1997); it is also that external and internal cognition are organised so as to work together for the achievement of cognitive goals. There is a well-designed division of cognitive labour. This is why the assertion that scientific laboratories are distributed cognitive systems is more that mere labelling: it implies that information flows within a system endowed with a cognitive architecture, and that the cognitive production of laboratories are the output of organised joint cognitive activity.

- *Lastly*, distributed cognitive systems have cognitive properties that differ from the properties of its constituent elements.

It is these last two points that justify the study of distributed cognitive systems *per se*, as entities whose study impart genuine and specific knowledge. The theoretical framework of distributed cognition not only leads to the recognition of the elements that participate to the cognitive processes, it also calls for the identification of the cognitive functions held by these elements, and their role in the achievement of larger cognitive goals. The analysis of the distribution of cognitive functions determines which element does what, for which purpose, with which inputs coming from where and which output directed at which other elements. The several elements that participate to the processing of information in distributed cognition are organised and form a larger computational system.

In managerial studies, the fact that different companies with similar characteristics (number of employees, techniques used, etc.) can have very different results is attributed to organisational factors, coined 'factors X' by Harvey Liebenstein. For cognitive production, this relates to the point that distributed cognitive systems have causally significant cognitive architectures. How cognitive labour is divided makes a difference. Moreover, distributed cognitive systems have emerging properties. Emergence characterises a certain type of relation of properties of complex whole with respect to properties of their parts, given that the emergent properties of the whole is possessed by none of its parts. Classical examples of emergent properties show how aggregated actions of agents give rise to a new property observable at the population level. For instance, a traffic jam can be seen as a higher entity which, in certain simple conditions where drivers slow down only when there is another car close ahead, is moving backward with regard to the direction of the cars. Clark (1997, pp. 73-78, 107-113), when analysing situated and distributed cognition, is especially interested in more complex phenomena of emergence, which involve heterogeneous agents and interactions with the environment. His examples are termites' construction of arches in their nests and Steels' robot. Termites make mud balls and add a chemical trace to them. When they deposit their mud ball, they choose the place where the chemical trace is strongest. The consequence is that they make columns, and that columns that are sufficiently close attract the deposit of mud balls that eventually join at the top. The second example makes clearer the role of the environment: Steels' robot finds its path towards a charging station indicated by a light source thanks to two behavioural systems leading to a zigzag approach to the light source and obstacle avoidance. The path of the robot, as successfully going to the charging station, emerges from continuous interaction with its environment. On the contrary, a non emergentist design for the path to the charging station would have the robot calculate its path in advance through *a priory* analysis of the constraints in the environment. In both the termites and the robot example, there is "a functionally valuable side effects brought about by the interaction of heterogeneous components". Clark underlines that well designed organisations do not necessarily result from the action of a central organiser; on the contrary, as Resnik (1994), he insists on the pervasiveness of complex emergent organisation with decentralised mindsets.

Hutchins (1995) has a similar argument about social organisation: "social organizational factors often produce group properties that differ considerably from the properties of individuals" (p. 175). It is a well-known fact that cooperating individuals can realise things, such as lifting a heavy stone, that non cooperating individuals are incapable of, regardless of the effort and time they spend on the task. This applies to cognition: "When the labor that is distributed is cognitive labor [...] the group performing the cognitive task may have properties that differ from the cognitive properties of any individual" (p. 176). The properties of a distributed cognitive system do not only depend on the properties of its elements, but also on how the cognitive tasks are distributed, i.e. how the system is organised. Thus "differences in the cognitive accomplishments of any two groups might depend entirely on differences in the social organization of distributed cognition and not at all on differences in the cognitive properties of individuals in the two groups" (p. 178). The functional design of social organisation can be accounted for thanks to the same 'emergentist' explanation expounded above. Hutchins uses the metaphor of the ant colony on the beach (an extension of Simon's metaphor) to show how the structure of the environment, as made of trails leading to food, emerges out of the ant's simple cognitive rules and actions on the environment (such as leaving a chemical marker where they pass). This emerging structure then provides the basis for efficient food gathering. The historical development of the social environment on the ship and procedures for navigation are paralleled with the ants' structuring of the environment. It is out of the historical developments, characterised as a "collection through time of partial solutions to frequently encountered problems" (1995, p. 168), that navigational computation is made effective.

9.2 Regulatory representations for distributed cognitive systems

James G. March (1988) has long argued that the design of an organisation is never the simple realisation of its leader's plans. So explaining this design cannot be done by appealing to its leader's representation of it. Yet, social agents can conceptualise macro-social entities, as well as their designs and function; and they do, more often than not, represent actual and desired designs of organisation. With this cognitive power, social agents can, sometimes, act on social systems with the intention of transforming them, which depends on their desires and beliefs. So, even if it probably never happens that an organisation is the pure fruit of some leader's decisions and actions, and that no organisation ever fully complies with its leader's intended design, there still remains an important causal role for these representations. So the design of an organisation is the result of external constraints, and designed-oriented intentional actions and other social actions with no such intentions. This is also what Hutchins illustrates in his example of "organisational learning" (1995, chap. 6). He recounts how a new distribution of cognitive labour arose in a situation of crisis where electrical failures affected the functioning of the gyrocompass. What happened is that the plotter made a series of decisions that organised cognitive labour; decisions were implemented and some of the implementations revealed themselves as infelicitous, non-adapted, and were selected out. The other members of the navigation team also took part in the re-designing of the organisation by negotiating their cognitive tasks. Human cognitive agents act in function of their local perspective, such as their limited access to the data, and then negotiate their part of cognitive labour with the other agents. Local designs and adaptive interactions among the subsystems form a process of organisational learning, which is situated in between blind emergent processes and classical global design. Deflationary theories with regard to the role of planning – coming either from situated cognition of from studies of social organisations – develop strong arguments showing the limits and locality of planning representations and their effects. However, there is in human affair, an indisputable causal role taken by representations of courses of action to be taken. Hutchins' described process of organisational learning appeals to representations of local and temporally bounded solutions for adapting to specific situations. But distributed cognitive systems must rely on representations that have a regulatory power by specifying how information must be processed in types of situations. Distributed cognitive systems, indeed, are meant to deal with types of problems, rather than with only one specific instance. They systematically deal with types of input and systematically produce types of output. In the case of navigation, the type of the input is visual landscape, and the type of the output is geographic position. This systematic treatment of types of problem is rendered possible because the navigation team knows what to do, and act in accordance with that knowledge. An apparently important representation that regulates the processes of the navigation system is provided by the 'Watch Standing Procedures', one of the ship's documents that "describe actions to be taken and equipments and techniques to be used" (Hutchins, 1995, p. 28). Such external documents acquire their regulatory effect by having mental versions of some of its parts as guide to behaviour – approximate and ephemeral as these mental versions may be. The navigator told Hutchins how important the 'Watch Standing Procedures' is when Hutchins said he wanted to know how navigation work was performed: the navigator referred him to the document and commented "It's all in here". Of course, other representations are also regulating the way the work is done. One could emphasize the role of cultural knowledge or know-how, for instance. It remains that these set of representations have an effect on how other representations are processed. They are, as Sperber (1996a, p. 29) coined them, "regulatory representations" [1].

Sperber's 'epidemiology of representation' is an analysis of the flow, transformation and eventual distribution of representations. The analysis takes into consideration representations that are mental states and representations that are public things. Mental representations take place within the mind of individuals when they think of something. Public representations are parts of the external environment that are intended to, and actually cause, the production of mental representations – utterances, written symbols and work of art are public representations. The flow and transformation of representations cut across the internal/external distinction, since mental (internal) representations cause the production of public (external) representations, via behaviour, that cause the production of mental representations, via interpretation, and so on. Representations are distributed in a population and in the environment as a consequence of their flow and transformation. Representations have causal histories that can involve different people and the environment.

The epidemiological approach is much akin to distributed cognition analysis, because it studies representations as present in an environment as well as in the brains, and the processes through which they transform. Hutchins and Sperber both go beyond the classical definition of culture in cognitive anthropology, "as that which need to be known in order to operate reasonably effectively in a specific human environment" (Bloch, 1991, p. 183). They both pay much attention to cognitive causal chains that span brains and environment.

[1] Regulatory representations are characterised by their regulatory effect rather than by their regulatory content. Some representations can have a regulatory content without having regulatory effect, if they are unsuccessful. If having a regulatory content is not sufficient for having a regulatory effect, I let open the question whether it is necessary.

Distributed cognition analysis and the epidemiology of representation have differed in their focus of investigation. Distributed cognition analysis has investigated the processes through which representations are produced and transformed within a community so as to fulfil some cognitive task; the epidemiology of representations has investigated the spread and distribution of representations within a population as cultural phenomena. The distributed cognitive analysis describes how the navigation team and its tools produce a representation of the location of the ship; the epidemiology of representations questions why charts and copies of the Watch Standing Procedures are found in every navy ship, why angular locations of landmark are produced again and again, why the navigator think of dots on the chart as location of the ship. I will argue that, on the one hand, the epidemiology of representation can explain why they are distributed cognitive systems and how they subsist in time, and on the other hand part of the epidemiological question—why are representations distributed the way they are?—is provided by distributed cognition analysis: these representations are produced again and again because they are part of a cognitive process that fulfils a cognitive function. The cognitive process is, as Hutchins shows, the culturally developed way a cognitive task is being fulfilled. However, the description of the cognitive task and how it is fulfilled does not provide the causes that explain why the cognitive task is tackled, and why it is tackled the way it is. The notion of distributed cognitive system is a concept for *functional description* of social phenomena and individual's behaviours, which are interpreted as fulfilling some cognitive functions. The epidemiology of representation, by contrast, is a framework for a *causal account* of these same behaviours and social phenomena: one look for the events and properties of the environment that are proximal causes of the production of representation. This section eventually attempts to relate Hutchins' type of functional explanations to Sperber's type of causal explanations.

Distributed cognitive systems produce a systematic distribution of representations, which participate to the fulfilment of the function of the system. What are the causes of this distribution of representations? The proximal causes of the production of some such representation are the input upon which cognition proceeds to produce the output representation. For instance, the proximal cause of reports of angular location is the actual positioning of the boat with regard to some landmark; the proximal cause of the plot on the chart representing the position of the boat is the report of angular location. But other representations account for the output representations: the sailors who report angular locations and the navigator have representations of what they have to do. These representations should enter a causal explanation of the distribution of representation. They are regulatory representations. Distributed cognitive systems always involve a distribution of regulatory representations, which regulates the functioning of the system and the consequent distribution of those representations that fulfil its function. Moreover, the distribution of regulatory representations must be regulated in some way, in order for distributed cognitive systems to exist through time. The 'Watch Standing Manual' must be aboard every navy ship, every navigator must know how to manipulate hoeys, etc. Continuing this reasoning, we could obtain an infinite hierarchy of representations, starting from the representations that fulfil the function, continuing with their regulatory representation, and the regulatory representations of the regulatory representations, and so on. This is not what happens in real life, if only because no infinity of representations are being realised in the world. Consequently, distributed cognitive systems must involve a *finite* set of representations whose distribution regulates the production and distribution of its own representations.

Sperber generally characterises institutional phenomena by hierarchical chains of mental and public (i.e. external) representations, causally linked, and where "the distribution of regulatory representations plays a causal role in the distribution of the other representations in the same complex." He then goes on to show how marriage, in France, complies with the characterisation. The French marriage institution involves two types of representations: first, representations such as utterances that declare a couple husband and wife, pronounced by a civil officer, second, regulatory representations of a courses of actions that describes the first type of representations and the conditions under they can be produced and distributed. For instance a chapter in the Civil Code is a public, external, regulatory representation. In a later chapter (p. 76), Sperber defines institutions as follow:

An institution is the distribution of a set of representations which is governed by representations belonging to the set itself.

Distributed cognitive systems involve regulatory representations that specify how information must be processed. The *Watch Standing Manual* is a regulatory public representation that is comparable to the *Civil Code* in its regulatory effects: it governs, to some extent, courses of actions, which determine the distribution of representations for navigation. In that sense, distributed cognitive systems are specific kinds of institutions. They are institutions endowed with cognitive functions.

Also, distributed cognitive systems, as institutions, are maintained across variably long periods of time. The cognition of distributed cognitive systems differs from the cognition of situated cognition in that the former tackle types of problems rather than just one token of a problem. In order to tackle types of problems, systems have to remain operative while new tokens of the same type are presented to them. The processes through which institutions last can be found in Sperber's definition, where the set of representations that constitute, together with their distribution, an institution, include a subset that regulate the production and distribution of the whole set of representations. In other words, institutions are, to some extent, self-regulatory. The 'governing' of the distribution of representations is done through causal chains that include mental events, ensuing behaviour and production of public representations that have further effects on people's mental states. Importantly, the regulatory representations are regulated by representations from the same set. So there must be some causal loop that goes from the effects of these representations to the cause of their production and distribution. It is this relatively autonomous regulation of the distribution of representations, Sperber says, that makes institutions self-perpetuating. There are, however, cases where the regulatory representations govern the extinction, rather than the maintenance, of the representations of their set. Secret communications, for instance, are set of representations with regulatory representations that strongly limit the distribution of the set. Most secrets are meant to die away with the people that know them. A more dramatic example is provided with small religious community with rituals and beliefs, among which is the instruction to commit suicide on a given date. Most probably, the beliefs and rituals will die with the community. The point of these examples is that self-perpetuating institutions must include a feed-back loop that maintain or increase the distribution of representations that make the institutions. Such feedback loop have been described by Barnes (1983) and Bloor (1997a).

Barnes (1983) and Bloor (1997a) citep [but see also][]searle95 describes a social process causing the perpetration of institutions, which also involves both intentional actions, i.e. actions whose causes include a mental representation of the goal of the action, and emergent properties. He takes the classical example of money: social agents attribute value to money and give it a function for the facilitation of the exchange of goods. With such representations, they use money in a way that *makes* it valuable and exchangeable. Through their actions, social agents are reinforcing the value and 'exchangeability' of money, i.e. the grounds of their own representations. Agents represent money as having a function for the exchange of goods, but the actual function of money is an emergent property of the aggregated actions of agents representing money as functional. Exchanges of goods mediated by money transfer, i.e. purchases, are social actions that have a positive feedback effects on the institution of money: it preserves, through time, its functionality by guarantying that it is, indeed, exchangeable. What we have, therefore, are social cognitive causal chains – purchases – which implicate the representation of money as exchangeable; this representation is maintained through the observation that there are purchases, which guaranties that money is exchangeable; and this representation generates further transactions. Barnes characterise this feedback process as partially self-referencing because the representation of money as exchangeable is rendered true by the fact that the representation itself is sufficiently spread across the population. This feedback process induces the perpetration and stability of the social causal chains of purchasing, through self-referencing representation, thus making a cultural phenomenon. Barnes and the Strong Programme additionally assert that this kind of feedback loop is pervasive in the maintenance of social institutions, including scientific and technological institutions. Does this self-referencing feedback process contribute to the existence of distributed cognitive systems? Distributed cognitive systems exist only to the extent that they are themselves endowed with some cognitive functions. People often have at least partial representations of distributed cognitive systems and their cognitive functions. These representations are regulatory representations, which can lead to self-perpetuating feedback loops. For instance, the captain of the ship thinks of the navigation team as performing an essential task for the overall functioning of the ship. Members of the navigation team also understand, to some extent, the encompassing cognitive task of navigating and the social organisation performing the task. These representations of the organisation of navigation contribute to fix the boundaries of the system (you belong to the navigation team or you don't), and target intentional feedback on the system. Positive feedback can include allocation of funding as providing wages for the team members and renewed material, and design oriented action, such as writing down successful procedures in manuals, thus maintaining the organisation of cognitive tasks and reasserting the status of the navigation team. Representations of distributed cognitive system, including expectations on their output and appeal to their results, can generate positive feedback action on distributed cognitive systems, thus contributing to the existence of the reference of the representation. This kind of feedback loop is the one described by Barnes.

Kurz-Milcke, Nersessian, and Newstetter (2004) argue that new members of the lab must to a certain measure learn how the lab functions, and that includes learning how it functioned in the past. Giere and Moffatt

(2003), writes that understanding the workings of the system is an element for evaluating the apparent veracity of the results. Sociologists of science would emphasise the role judgments of funding bodies have upon the functioning and results of funded laboratories. Representations of scientific institutions play a role as regulating the cognitive processes through which results are produced, and they play a role in originating positive feedback loop.

The explanation of rational design as emerging from collective action rather than from a single agent having a rational representation of the design suits well social epistemologists and sociologists of scientific knowledge, who have long argued that the achievements of science depend on the social aspects of scientific practices. The explanation of rational actions from the viewpoint of situated and distributed cognition is twofold: first, rational action is understood as stemming from the interaction with the environment; second, the environment is itself the product of collective social actions. The advent and evolution of scientific thinking, in particular, is not due to change in the human mental apparatus, but in the environmental cultural conditions of cognition. Moreover, these conditions are not the product of one great thinker – as, say, Galileo providing the ultimate cognitive tools and methods for science to develop – but emerge through the actions of multiple agents. This points out an important similarity between scientists' rationality and the ants' rationality: they are both the result of the interaction of some organism pre-equipped for social interactions (the ant is 'communicating' by secreting, and favouring path with, pheromone) and a constructed environment.

There are nonetheless important dissimilarities between the social organisations of non-human animals and humans. One of the differences is that human social organisations evolve independently of genetic change, while the basic patterns of other animals' social organisations seem invariable within species. Insisting on the role of representations of social entities and their properties in the historical construction of social organisation is important because such representations are most probably human specific, while social organisation, loosely specified, is not. Resnik (1994), for instance, gives numerous examples of animal, yet highly complex, social organisations, which emerge out of the aggregation of actions that stem from simple mechanisms. Considering the role of human specific representations in framing social organisation may provide an important insight on human institutions.

Unfortunately, the demise of the central organiser has led authors such as Clark and Hutchins to pay little attention to the role of actions based on representations of organisations or institutions. But design-oriented actions can also have effects even when the fit is not perfect between the state intended as a result of the action and the actual resulting state. Moreover, representations of social entities can play a part in the historical construction and maintenance of the social entities. Money is a case in point. Hutchins' analysis of the history of the Western navigation system shows the respective role of representations of local solutions and the historical processes out of which systematic treatments emerge. He says: "We are all cognitive bricoleurs – opportunistic assemblers of functional systems composed of internal and external structures". Hutchins' book tells us what sorts of cognitive bricolage are done and contrasts it with classical internalist cognitive science. Once the role of external structures is settled, one may ask what role remains for the internal structure. Often, the answer takes the form of an argument in favour of an understanding of the mind as a pattern matcher. Another less addressed question, however, concern the cognitive processes through which external structures are made to participate to cognition, how assembling of functional system is done. The questions I will address are 'what is it that makes us cognitive bricoleurs?', 'what are the representations involved in cognitive bricolage?', 'what are the cognitive processes that make cognitive bricolage?' I will argue that human mental representations of entities as cognitive functional entities is exactly what sustains human distributed cognitive systems and sets their evolution in motion. I will develop this argument in the next chapter (chap. 10). But first, I clarify below the notion of function of distributed cognitive systems.

9.3 Functional analysis of cognitive systems

At this point, it is useful to emphasise a difference between Hutchins' framework, and the framework of situated cognition (e.g. Clark, 1997). While the latter is an investigation of how human beings exploits their environment, acts upon it and take advantage of its regularities for cognitive processes, the former goes one more step away from the individual isolated cogniser and towards macro-sociology. While situated cognition still pertains to cognitive psychology, Hutchins' notion of distributed cognitive system contributes to the social sciences. Hutchins makes a jump from micro-social analysis, concerned with socially situated and interacting individuals, to macro-social analysis, concerned with social organisations and institutions. At the macro-level, social organisations can be described as performing cognitive tasks through the good functioning of its elements. But what justifies this attribution of cognitive properties to social entities?

What is apparent to the cognitive ethnographer are, first and foremost, people using artefacts and communicating. The interpretation of these events as cognitive leads to understand the informational load of the events observed. Information bearers, i.e. representations, are manipulated, transformed, communicated. Their relations and effects through the network of individuals constitute social cognitive causal chains, i.e. causal chains made up of entities with semantic properties, which span through social agents via their productions of perceptible behaviour (e.g. utterance) and effects of behaviour (e.g. written symbols) intended to cause mental representations in others (Sperber, 2001b). However, the ethnographer can soon observe that the social cognitive causal chains are reproduced again and again. Such reproduced social cognitive causal chains are called by Sperber 'Cultural Cognitive Causal Chains', because they cause the lasting production of versions of representations, thus stabilising them in time and within a community. For instance, the 'fix cycle' implemented by the navigation team when coming to port, is a social cognitive causal chain that is reproduced every 3 to 7 minutes and that invariably leads to position fixing. Fix cycles produce external representations of the localisation of the ship through a process involving observing the visual bearings of the landmarks, communicating the results, drawing 'lines of position', etc. What is the reason why these social cognitive causal chains get reproduced again and again? Why are the representations involved stabilised? Sperber's epidemiology of representation is a research programme that aims in particular at answering such questions. It aims at understanding why representations stabilise among a community, and through which cultural cognitive causal chains. A reading of Hutchins' cognitive analysis provides one specific answer to Sperber's question: some social cognitive causal chain get reproduced because they implement a cognitive function. When being reproduced, they also stabilise sets of representations. The reproduction of these social cognitive causal chains is caused by higher-level representations that regulate their production. But how and why are these regulating representations stabilised? I argue that it is because of their determining role in implementing the cognitive function.

The functionalism present in the notion of distributed cognitive system sends us back to two relatively disjoint theories: functionalism in cognitive science, as grounding the semantic description of physical events, thus clarifying the ontology of cognitive processes; and functionalism in sociology and anthropology, as a means to account for what macro-social entities do. These two kinds of functionalisms are relevant because distributed cognitive systems are both macro-social and cognitive entities. While the theories have had independent scientific histories, they both stem from a first observation that the social or the mental world is well designed; it is designed for a purpose, and to achieve specific goals. According to the functionalism of Malinowski social entities such as institutions are to be understood as serving the biological needs of individuals. According to the structuro-functionalism of Radcliffe-Brown or Talcott Parsons social entities serve the needs of the society at large. According to functionalism in the philosophy of mind and cognitive science, the biology of the brain, just as that of any evolved organ is to be understood as contributing in a specific way – in this case by guiding its behaviour – to maximising the "fitness" or reproductive success of its possessor. The consequence is that it is possible and fruitful to theorise on what the brain does independently of how the brain does it. What the brain does is give instructions to the body depending on the environment. We thus have an input/output device (corresponding to sense/behaviour) that can be described as processing information. As a consequence, cognitive psychology can start up studying what the brain does with a relative independence from advances in the biology of the brain ("the mind is what the brain does", says Dennett).

A common trait of social and cognitive functionalism is that they justify a new level of description of the phenomena: cognition as accounting for what the neurons of the brain do, and macro-social functional analysis as accounting for what people in social institutions do. However, cognitive functionalism restrains functional analysis to one particular mode of performing a function, which is to process information. Hutchins non-explicitly proposes to apply functionalism to the analysis of social phenomena, and then restrains his analysis to cognitive functions of social entities. The functionalist choice that sustains Hutchins' theory of distributed cognition raises questions. Both social and cognitive functionalisms have met with difficulties and sceptical arguments [2]. Interestingly, Hutchins claims to save cognitive functionalism, a theory developed in

[2] I will not review the criticism against cognitive functionalism (but see, e.g. Block, 1980, 1994), for I think the theory is true and remains at the centre of cognitive science, and that the criticism is not hitting the target. I just want to mention one difficulty with cognitive functional analysis: information on input and output largely underdetermines psychological theories about the cognitive processes at work. The problem, however, finds its solution with interdisciplinary work on the neural implementation of cognition and on the biological evolution of the cognitive apparatus (the brain) (Mundale and Bechtel, 1996). The latter leads to evolutionary psychology, which provides a powerful heuristic tool for the investigation of cognitive functions, seen as produced by evolutionary history. This same investigative strategy, which calls on history and micro-analysis (at the level of neurons) to understand functions, is similarly worthy in the social sciences.

the 60′ and which is still at the heart of cognitive science, with the implicit use of social functionalism, a theory of the first half of the twentieth century that anthropologists no longer openly endorse. Yet, functionalism still provides an important explanatory resource. Within an epidemiological perspective, I want to argue that functions describe some specific social causal chains which include feedback processes. In order to specify an appropriate functionalism for the study of socially distributed cognition, I consider and answer some arguments advanced against it.

With regard to functionalism in the social sciences, it has been emphasised that most social entities (or at least some of them) do not serve a function that is beneficial to the individuals of the community (social phenomena include symbolic structures with no beneficial social function; most social entities do not maximise social benefits and would appear as irrational and unaccountable in a functionalist framework). This type of argument shows that a functionalism that would restrain sociological explanation to functional explanation is erroneous. But it leaves the possibility open that *some* social phenomena get explained by their function. So let us just admit that Hutchins does not intend to explain all cultural events with the notion of cognitive function. In fact the distributed cognitive framework is a non-reductive tool opening up new possible explanations in terms of cognitive functions, and should be used within some larger framework of non-reductivist functionalism in the social science.

Another criticism against functionalism in the social sciences bears on its inability to account for cultural change. One of the solutions, however, is to integrate the role of the environment. The idea is that social systems are themselves situated. So the beneficial properties of their function are dependant on the properties of the natural and social environment. For instance, an institution whose function is to re-distribute a certain type of goods looses its beneficial properties if the goods come to disappear. An institution that has for function to insure good relations between two kinds of social classes looses its beneficial properties if the social classes go extinct. Factors of change are often attributed to technical innovations. A non-reductivist functionalism allows factors of social change to come from elsewhere, such as from the emergent consequences of individuals' aggregated actions. As opposed to the structuro-functionalist, which describes societies as closed systems of functional social entities that maintain a self-reproducing equilibrium, a non reductionist functionalism can place functional social entities in a changing world. Talking of evolving distributed cognitive systems is already talking about changes in the social organisation of the systems. But, although evolution was not primarily integrated in social functionalism, the notion of evolving system is not paradoxical. Hutchins thus talks about the evolution of the western navigation system, which was generated, he suggests, by adaptive responses to new situations. The new situations he mentions are especially those involving technical innovations and inventions of new representational media. A technical innovation can lead to the full reorganisation of computation even when it initially concerns only a small aspect of the computation. This is because of the computational ecology of tools, where each tool is a part of the computational environment of the other tools. Hutchins (1995, p. 112-114) provides a set of examples where some technological innovation are useless for navigation ... until some other innovation renders it exploitable.

The mutual dependencies among the various instruments and techniques is clearly visible in the history of navigation. Even though the chip log was available for use in the sixteenth century, for example, it was not generally adopted until the middle of the seventeenth. Why weren't sailors using the log more widely? Because they had no convenient way to carry out the computation required to turn the readings gained from the log into useful information about the ship's position. (p. 112)

Finally, the most fundamental criticism lies in the nature of functionalist explanations: functionalism explains the existence of social entities by their beneficial effect on individuals and society. But effects are not explanation of existence. A genuine explanation describes the causes out of which the social entities came into existence. The answer to this criticism is to show that the effects of some macro-social entities do have a causal impact on their existence through the existence of feed-back loops. The most simple feedback loop is selection: either the effect is satisfactory and the social entity is preserved, or the effect is not satisfactory, and the entity is selected out. Selectionist theories in the social sciences have more recently been applied to the study of the development of social organisations (Nelson and Winter, 1982). Depending on the school of thought, selection is said to operate on organizations, organisational practices or routines, and on resources as ability to product specific results (e.g. if an enterprise is not sufficiently efficient, then it closes down; or controlling end products, as a routine, can be selected out in favour of controlling products at all levels of the production chain). A key point of such feedback loop, in any cases, is that their causal effects must be consequential on macro-social entities. The feedback consequence on macro-social entities is often taken to be some emergent feature stemming from multi-causal events happening at the micro-level (see the previous section on emergence). In the case of evolutionary biology, as a paradigmatic case of causal account sustaining functional analysis, species and organs are selected out through the numerous events of natural selection that span over relatively large period. Selection of species emerges out of multiple events related to survival and reproduction. For social entities, Barnes (1983) and Bloor (1997a) have described the feedback loops involved in the preservation of institutions and organisations (see also the previous section).

9.4 DISTRIBUTED COGNITIVE SYSTEMS AS INSTITUTIONS ENDOWED WITH SOME COGNITIVE FUNCTIONS

Distributed cognition, as a conceptual tool for the social sciences and for science studies, comes with functionalism, since distributed cognitive systems and their internal components are characterised by their cognitive functions. The notion of function is in fact already present in the observation that social organisations can have a design, because 'design' in this context means organisation for the achievement of some attributed goals, or, in other words, well-organised for performing its functions. Functional analysis can be limited to the description of the function of some entities.

Or it can enter the details of how the function is performed by looking at what the entities actually do and how they do it, as for instance how the tasks are distributed among the components of the entities. This is describing the design of the entity, which is what most studies of distributed cognition do. But a comprehensive and naturalistic functional analysis includes the causal history which accounts for the emergence of the function. Such an analysis includes the design-oriented actions of agents and/or the feedback process through which the functions of the entities contributed to their existence and stability.

To specify a cognitive function is to describe some social cognitive causal chains that are reproduced and maintained across time and which have semantic properties that serve some cognitive goal. The goal is a goal to the extent that achieving it contributes to the reproduction and maintenance of the social distributed causal chain which achieved it. So there must be a feedback loop for the effect of the social cognitive causal chain (the result of the function) to have a consequence on whether the social cognitive causal chain shall be reproduced or not. This can happen, in particular, if the goal and the system are represented by agents that have the power to act upon it.

The analysis of distributed cognition requires identifying a cognitive goal and the system that aims at achieving it. Since the boundaries of cognitive systems are not any more fixed by the skin of individuals – as Hutchins puts it – they must be determined on other grounds. Circumscribing cognitive systems may prove to be controversial. Giere (2003) points this problem out and calls it, following Clark (1997), the 'identification problem'. He seems to suggest, for instance, that the objects under scrutiny by scientists enter the cognitive system (in his example, a cluster of galaxies) [3], while the generating plant that provides the necessary elec-

[3] He says: "Additionally, Hutchins' system is confined to the deck and pilothouse of a ship. The Hubble Telescope system [as the distributed cognitive system studied] extends at least from Earth orbit to the State of Maryland. If we count the lensing by the Abell 1689 galactic cluster as part of the system detecting the distant galaxies, the cognitive sys-

tricity does not. He then concludes that there seems to be "in practice little difficulty in deciding whether particular things are or are not playing a relevant causal role in the process of achieving the cognitive goal" (Giere, 2003). I agree with Giere that circumscribing a cognitive system is a question that must be answered *a posteriori* and that may not require a criterion made of necessary and sufficient conditions. On the other hand, the circumscription of a cognitive system cannot only be decided with pragmatic considerations that serve the scientist's purpose only. For instance, the cluster of galaxies rather than figuring in the distributed cognitive system, belongs to the environment of the system that gives it input. How are we to settle our disagreement? The question is not futile because if cognitive distributed systems are to qualify as actual systems, then their limits need be settled on empirical ground related to the organisation of the cognitive phenomena. It must be possible to achieve a good balance by avoiding the pitfalls of undue reification of cognitive systems, but not missing the phenomena that constitute them as relatively autonomous cognitive entities. There are several constraints that enable circumscribing distributed cognitive systems. The method used by Hutchins consists in identifying a cognitive task, then finding out the elements that perform the task. How to identify a cognitive task that is actually tackled by some cognitive distributed system? How to circumscribe which elements have a causal role in performing the task, rather than being part of the environment of the cognitive system? I argued that cognitive analysis always goes together with a functional analysis. An account of an evolving distributed cognitive system cannot be restricted to the faithful description of the process-

tem is distributed 2.2 light years out into space. [. . .] What counts as part of the cognitive system? Although this may seem like a serious problem if one insists on explicit criteria, there seems to me in practice little difficulty in deciding whether particular things are or are not playing a relevant causal role in the process of achieving the cognitive goal. It is abundantly clear that the Abel 1689 cluster played an important causal role in producing the final image and the conclusions drawn from it. It must therefore be included in any account of this cognitive achievement." While Abel 1689 may be included in an account of the cognitive achievement, it is not clear whether Giere thinks that it should be included as an element of the system. Note also that one can choose to put the existence of Abel 1689 into parenthesis and refer only to the input on the basis of which its existence is asserted ing of information that is done. The processing is functionally described as aiming to attain cognitive goals, which are themselves understood with the stakes set by the situation. For instance, navigation as cognition goes together with the knowledge that bad navigation can lead to wreaking the ship, which is an outcome to avoid. Bringing up functional analysis in the study of distributed cognitive system furnishes genuine criteria for finding out their boundaries: these are set by the scope of the feedback loops. In particular, when people have a representation of the distributed cognitive systems and act upon them in order to build the system they think of, then it is safe to assert that the distributed cognitive system is the one that is represented by the agents. One problem of specifying the boundaries of distributed cognitive system comes for the facts that people may not have representations that specify clearly which element is in the system and which element is not, and there may be a distribution of different representations about the elements of the system. The systematicity of distributed cognitive systems need not be strong, it depends on the distribution of representations that regulate it. Contrary to, say, biological organism, distributed cognitive systems can have little specified boundaries. However, the conditions for circumscribing a set of interacting elements and qualifying it as being a distributed cognitive system include: the set of interacting element has been attributed a cognitive goal; achieving its goal is its function; it has a cognitive architecture, which appears designed for performing its function; the systems of interacting elements is preserved only if it performs its function (feedback loop). In brief, identifying a distributed cognitive system is identifying an institution, as characterised by Sperber, which has a function, as characterised by the functionalism in social science described above, that is cognitive, which means that the description of the function of the institution is best done in terms of flow of information or transformation of representations.

A consequence I have pointed out is that functional analysis is essentially historical. This point was long made in biology, with Darwinism as explaining the design of functional organs. In the social sciences two, design is best explained by the historical conditions into which it appeared and developed. Consequently, a full understanding of distributed cognitive systems should include an account of their evolution.

Chapter 10

The evolution of cognitive systems of science

This chapter is devoted to the study the mechanisms through which distributed cognitive systems evolve. The first section shows the relevance of such a study for the history of science. In the sections that follow, I attempt to describe the psychological and environmental factors of the evolution of distributed cognitive systems. I argue that humans entertain and distribute representations about the trustworthiness of people and things, and that it is the distribution of these representations that regulate the evolution of distributed cognitive systems.

10.1 History of science and the transformation of distributed cognitive systems

The main idea of this section is that the notions of Distributed Cognition can fruitfully be used for the history of science, but that the framework needs to be complemented in order to account for the evolution of cognitive system. If, at any one time, the organisation of scientific knowledge production can be understood as constituting distributed cognitive systems, then historians should wonder how one passes from one cognitive system to another. Thagard, Giere and Nersessian have argued that the practice of science constitute and takes place within distributed cognitive systems. These arguments and case studies should motivate us to go ahead with the distributed cognition framework in the domain of history. In fact, Giere has already offered some insights on the topic. With regard to the introduction of new external representations and new artefact in science, Giere and Moffatt (2003) (2003) say:

The invention of new forms of external representation and of new instruments for producing various kinds of representations has played, and continues to play, a large role in the development of the sciences. From a cognitive science perspective, both sorts of invention amount to the *creation of new types of distributed cognitive system*.(My italics)

In another article, Giere (2002a) even adds an explicit note about the historiography of science. Historical studies of science, he says, should see the scientific revolution as the creation of new distributed cognitive systems. The relevant section of the article is called 'Distributed cognition in the history of science':

It is often claimed that the scientific revolution introduced a new way of thinking about the world, but there is less agreement as to what constituted the 'new way'. The historiography of the scientific revolution has long included both theoretical and experimental bents. Those on the theoretical side emphasize the role of mathematics, Platonic idealization, and thought experiments. The experimentalists emphasize the role of experimental methods and new instruments such as the telescope and microscope. Everyone acknowledges, of course, that both theory and experiment were crucial, but these remain a happy conjunction. The concept of distributed cognition provides a unified way of understanding what was new in the way of thinking. It was the creation of new distributed cognitive systems. Cartesian coordinates and the calculus, for example, provided a wealth of new external representations that could be manipulated to good advantage. And the new instruments such as the telescope and microscope made possible the creation of extended cognitive systems for acquiring new empirical knowledge of the material world. From this perspective, *what powered the scientific revolution was an explosion of new forms of distributed cognitive systems*. There remains, of course, the historical question of how all these new forms of cognitive systems happened to come together when they did, but understanding the source of their power should now be much easier. (My italics)

This suits well the case study developed in the next chapter - the introduction of computers in the practice of Mathematics -, which also amounts to the creation of a new type of cognitive system. The introduction of computers in Mathematics is therefore comparable to the introduction of other artefacts in science such as the microscope and also to the introduction of new external representations in Mathematics such as Cartesian coordinates.

Nersessian and her colleagues have also grasped the importance of the history of distributed cognitive systems for the history of science. They study distributed cognitive systems as evolving cognitive systems, using a "mixed-method approach" for the incorporation of cognitive-historical analyses. Their historical analysis includes a description of the processes through which a distributed cognitive system is maintained through time in spite of the fact that its elements are changed because of turnover (the members of the lab where they did their fieldwork are mostly on short term contracts). They also convey the temporality of the models of their lab – mental or material simulations of natural phenomena. They describe, for instance, how new students or post-doc with new projects make new use of old artefacts (Kurz-Milcke et al., 2004). The pervasive presence of the models in the lab and their relatively rapid change through interactions leads the author to talk about the "fabric of interlocking models".

Latour has shown that much of what is at stake in the History of Science is the way new means of knowledge production are recruited by and for science. Giere translates this insight and shows how the history of science is in great part a history of distributed cognitive systems for knowledge production. With scientific development, there is an evolution of cognition, but this evolution is realised in the socio-cultural structure. This implies that a key process in the evolution of science is the process through which distributed cognitive systems evolve. So a mixed-method approach where cognitive history is history about distributed cognitive systems may be the appropriate approach. In the previous section, I have characterised distributed cognitive systems as institutions endowed with a cognitive function. Consequently, the problem of the evolution of scientific cognition is foremost a sociological question about the evolution of institutions. Within an integrated perspective, however, one also has to ask what the mental cognitive processes that enable such evolutions are. It is by going back and forth from the mental to the social implementation of cognitive processes that a mixed-method approach will prove most useful, and lead to a comprehensive understanding of the causes of changes in the organisation of scientific labour.

10.2 How cognition culturally evolves

Cognition evolves with socio-historical development. This assertion stems from the unrestricted notion of cognition, which includes non mental processes and thus leads us to considerate seriously the changes that take place, over time, in socially distributed cognitive systems. The evolution of distributed cognitive systems is an evolution of types of cognitive processes implemented in social organisation. Therefore, it is cognitive evolution. How does this evolution proceed ?

It may be necessary to note that the concept of evolution is not necessarily a teleological one. It needs not either refer to progressive processes, where present states are judged to be 'more something' (better, more complex, more satisfactory, more efficient or more advanced, for instance) than past states (e.g. cumulative evolution). 'Evolution' refers to temporal processes where earlier states play a crucial role in the determination of later states. 'Evolution' in evolutionary biology or in epidemiological studies, for instance, refers to the changes in the frequencies of types in a population.

Although Hutchins dedicates some parts of his book (1995) to the historicity of distributed cognitive systems, the research tradition working on distributed cognition has, as far as I can tell, devoted little effort on the analysis of the historical evolution of cognitive systems, strongly privileging functional-structural analyses describing how cognition is distributed and investigating the rationale of the distribution, rather than investigating how cognition happens to be distributed the way it is. The latter question is however important, if only because the analysis of cognitive systems, as functional systems, is founded on principles of evolution.

An independent and older research trend, however, has investigated the evolution of cognition as a consequence of the use of a new representational medium – written symbols. The research includes the work of Goody (1977), Olson (1994), and Donald (1991)[1]. These authors have described in details the evolution of cognition induced by literacy. From a distributed cognitive perspective, literacy consists in attributing cognitive functions to external objects that are endowed with symbolic properties. For instance, writing provides an external memory system with processes for the conservation of information and for information retrieval that are different from those of mental memory. Goody (1977) also includes analyses of the usage of symbolic external representation such as grids or tables, which changed individual practices, cognitive possibilities, and cognition. For Merlin Donald, symbolic literacy consists in a third stage in the co-evolution of culture and cognition – the two first stage being implemented through changes in the human brain. The third stage, Donald says, corresponds to the "externalisation of memory". These studies of the cultural and historical evolution of cognition emphasise the role of artefacts as cognitively potent through symbolic interpretations. They therefore contribute to the study of distributed cognition, enriching it with historical studies and with the evolutionist framework they bring in.

The evolution of cognition also importantly involves the acquisition of new knowledge. This is, in particular, the traditional view of the evolution of scientific cognition as an incremental accumulation of scientific knowledge. Such a view makes little room for the social and distributed aspects of scientific cognition. The notion of distributed cognition enables describing the role that newly acquired knowledge may take in the production of

[1] These authors are surprisingly not quoted by Hutchins (1995) and Clark (1997), with the exception of one quote of Goody (1977) from Hutchins only to mention that his 'cognition in the wild' does not refer to the distinction between wild and domesticated mind alluded to by Goody.

further knowledge. Knowledge production renews its means of production with its own production. For instance, I have suggested that scientific paradigms could be seen as distributed cognitive systems that give a special role to some theoretical claims.

Thirdly, the evolution of cognition can be the result of change in institutions and social organisation. Institutions distribute and constrain cognition, and specify cognitive tasks. The role of scientific institutions in the development of science have not been studied in a cognitive perspective, but they have been an important object of study for sociologists of science – which saw them as constraining and empowering scientific thinking (Bloor, 1997a). Institutions are defined as shared rules supported by various enforcement mechanisms. So, in this perspective, the social organisation of cognitive labour is not a focus of research, hence the essential import of the distributed cognition framework. Within this framework the social organisation of cognitive labour relates to the flow of information within a social structure, and this structure can include artefacts as well as people.

10.3 MINDS AND THINGS IN THE MAKING OF DISTRIBUTED COGNITIVE SYSTEMS

I have noted three 'socio-cultural means' by which cognition can evolve: use of new artefacts, use of new knowledge, and change in social organisation/cognitive architecture. In fact, these three means are strongly interrelated: new knowledge and new artefacts are given a role in the historical development of science only if they enter the constitution of new distributed cognitive systems. This involves changing the cognitive architectures of social organisations. Changes in social organisation often require, and are often initiated by, the acquisition of new knowledge – as for instance technological knowledge that leads to the ascription of cognitive functions to the new technique. Using new artefacts requires one to know how to use these artefacts. What is the primum movens that drives these changes? Some think environmental conditions are responsible for changes in societies and cultures, some say human decisions are.

Individualistic science studies may be tempted to find the primary cause in some of the ideas held by prominent scientists. Such prominent scientists, indeed, have introduced representations which, directly or not, are used as regulatory representations. They have henceforth created distributed cognitive systems. Thus Descartes has created a new distributed cognitive system for producing mathematical knowledge by introducing Cartesian coordinates; thus Galileo systematised the use of the telescope for astronomical observations, and thereby changed the distributed cognitive system for astronomical knowledge production. However, sociologists of science will promptly remark that the representations produced by prominent scientists acquire their regulatory power only if they are given this regulatory status by the scientific community. Ideas of scientists must be made, through social interactions, norms of good scientific practice and reasoning. It is only under these social conditions that distributed cognitive systems are created. Moreover, the thoughts and actions of individual scientists are better understood within the context in which they occur, which include the social and cultural situation.

According to technological determinism, a trend in Marxist theories, technology is the prime mover of social change. "The historic mode of production, i.e. the form of society, is determined by the development of the productive forces, i.e. the development of technology", says Bukharin (quoted by MacKenzie, 1996, p. 24). When applied to change in distributed cognitive systems, this means that the form of these systems, the social organisation, is determined by the available technology. This makes sense: the distribution of cognition is obviously constrained by the nature of the elements that form part of the distributed cognitive system. What are these elements? What can they do? Answering these questions is already understanding much of the causes of change in the organisation of distributed cognition. With the introduction of new techniques, some tasks and jobs disappear, because they become useless or unproductive, and some new tasks and jobs appear, such as the expert handling of new technical objects. MacKenzie (1996, chap. 2), following Marxist authors, provides a more complex picture of Marx's understanding of social change. MacKenzie wants to show that in spite of Marx's theoretical statements in favour of one-way determinism from technology to superstructures, via modes of production, Marx's historical analyses include statements about the determination of technological evolution by the social conditions of innovation. MacKenzie's argument is that, when confronting detailed historical data, Marx is lead to take into account the causal role of motivated actions for technological change. Studying the causal role of the social interests of those who actually have the power to influence the design of new technologies is at the heart of current research programme in sociology of technology. MacKenzie develops an argument in favour of this research programme by showing that the alternative account, technological determinism, does not stand face to the historical analyses of Marx, its best proponent. MacKenzie quotes Marx saying: "Man not only effects a change of form in the materials of nature; he also realizes his own purpose in those materials" (p. 26). MacKenzie then comments: "the inclusion of labor power as a force of production thus admits conscious human agency as a determinant of history: it is people, as much as or more than the machine, that make history". Marx's historical materialism would take into account the purposes of the people involved. Determination does not simply go from technology to society (its structure and ideology), but also from society to technology, via innovation. And it is through his purposes that Man orients technological innovation. But what are these purposes? MacKenzie goes back to Marx's historical analyses to point out the role of the "valorisation process". The valorisation process includes of course, profit maximizing, as with the capitalist bourgeois accumulating capital. But some pieces of Marx's historical analysis indicate that valorisation will also involve other interests, such as the creation and the maintenance of desired social relations (MacKenzie, 1996, p. 45).

Departing from simplistic pictures where distributed cognitive systems of science are merely designed by prominent scientists, or where their design is the mere consequence of an independently developing technology, we find again the importance of the relations between the mental and its environment, and the processes of change including both planned actions and emergent properties. Questioning the determination of 'good design' of distributed cognitive system with sociology and psychology shall lead to a comprehensive naturalistic answer. The questioning starts with a functionalist analysis, which 'naturalises' the characterisation of 'well designed' and draw attention to the historical construction and maintenance of distributed cognitive systems. These processes take the form of distribution of representations across interrelated social cognitive causal chains. The determinants of the distribution include:

1. the production of representations by scientists and technological innovators and the causal regulatory effects that some of these representations acquire;

2. the regulatory representations as specifying the processes of a cognitive system, and the representations regulated whose distribution implement the cognitive system;

3. the feedback loops, as causal chains that start with the production of distributed cognitive systems (the representations output of the systems) and that eventually determine the distribution of regulatory representations;

4. the conditions of the production of these representations;

5. the motives of the people involved in the chains;

6. the mental cognitive processes that determine their involvement in the chains;

The analysis of the first point amounts to a description of innovations and the events that lead to it. The description prepares the analysis of the causal determination of scientific innovation: the cognitive processes which lead to the innovation, including the distribution of these processes and the mental processes of leading scientists; the reception of innovatory thoughts by a scientific community – which is itself determined by the content of the innovation and by the social and cognitive conditions of its reception. What where the motives of the people who spread the innovative representations and implemented their regulatory content? What mental processes enabled the acquisition and transmission of innovative knowledge?

The second point calls for a description of the flow of information in distributed cognitive systems. If certain representations are regulatory, one must analyse their regulatory effects. This is the traditional analysis made in the framework of distributed cognition, which answers the question of which cognitive processes are implemented and how these processes are distributed among cognitive elements.

The third point makes clear the social cognitive chains out of which some institutions qualify as distributed cognitive systems, and thus as functional: the spread of regulatory representations is dependant on the satisfaction that sufficiently influential people derive from the output representations of the systems.

The fourth point recalls that elements belonging to the environment (outside the brain) participate to the cognitive processes implemented by the distributed cognitive system. The environment also includes the material conditions, which can play a role in the evolution of organisations and technology. The analysis of the fourth point calls for a description of distributed cognitive systems, as providing the context that is of special importance for innovation and for the implementation of associated organisational change. In particular, the social reception of new technologies depends on how useful the technologies can be made to attain the cognitive goals of existing distributed cognitive systems. Hutchins' remarks on the ecology of cognitive tools are fine illustrations of the importance of distributed cognitive systems as context of reception of innovations. The conditions of receptions play a role not only on the selection of competing technologies, but also in the creative act of designing technologies, since the designers are interested in the reception of the technologies they develop. The broader social context may also be relevant, because goals of existing distributed cognitive systems can be changed, and adapted to the needs and the means of the situated community. For instance, I will show that the integration of computers in mathematical practices, especially computers as taking an active part in the production of proofs, has slightly displaced the goal of mathematics. Why has this displacement been initiated and accepted? Distributed cognitive systems can evolve so as to incorporate new technologies, but change in the systems can be costly, generate conflict of interest, and require much effort for the stabilisation of the new systems.

MacKenzie's analysis of Marx shows the importance of the motives, or interests, of the people involved: why do they promulgate some representations and some their associated institutions? Why do they invent the things they do? Why do they participate in cognitive systems? All these actions require time and effort, which are scarce resources, and such resources are not usually spent without motivation. MacKenzie's analysis shows that the interest theory of the Strong Programme is already lurking in Marx's writing, and thus asserts the relevance of interest theory for the analysis of the relations between the history of social organisations and the history of technology.

For our purpose, the relevant relations are between the history of distributed cognitive systems and the advent of new cognitive elements. What MacKenzie illustrate, with Marx, is that while the advent of cognitive elements – as changes in the environment – do play a key role in the design of distributed cognitive system, the converse also holds: the actual distribution of labour determine technological and scientific innovations. The strength of the argument is that the cases in its favour are found in the historical analyses of Marx – who would normally be thought as an opponent of theories asserting social interests-to-technology determination. The division of labour determine technological and scientific innovations because:

1. innovations are the output of distributed cognitive labour;

2. social organisation determines people's interests, which determine

their actions (innovations included).

Saying that science is produced by distributed cognitive systems implies recognising that scientific institutions are 'well designed'. But such recognition raises further questions: 'designed for what?', and 'how is the design achieved?' As functional analysis shows, answers like 'distributed cognitive systems are designed for the production of truth', and 'their design is achieved through rational decisions' are not satisfactory. The sociological investigation of functional social systems requires investigating the actual social processes through which design is achieved. It is produced by intentional actions (i.e. actions with some goals, where the acting agent is interested in the consequences) and negotiated decisions, which yield emergent properties and feedback loops. Seen in this light, the analysis of the evolution of distributed cognitive systems sends back to a basic question in history: *Why* did the people involved do what they did? In our causal perspective, the intentions or the interests of the people involved are mental representations of a specific kind, which are causally connected to planning as the production of representations for action.

There remain fundamental processes not yet investigated: the cognitive processes through which some representations have regulatory effects. This refers back to human cognitive abilities that are not yet well understood in current cognitive psychology. In particular, regulatory effects are obtained when people understand the regulatory representations as having normative content. But how is this done? When we asked why people comply with the normative content of regulatory representations, we assumed that they do cognize norms as such. With the constitution of distributed cognitive systems, one kind of regulatory representation is fundamental: it is the sort of representations that ascribe cognitive functions to elements of the system. Such representations regulate the system in that they designate who, or what, does what. For instance, the *Watch Standing Manual* for navigation indicates which element, tool or person, has which function. It explains that the fathometer measures depth of the water under the ship, that the navigation plotter plots the location of the ship on the charts, etc. Such representations are also fundamental for changing distributed cognitive systems: a change in the distribution of cognition is the result of a new ascription of cognitive function. Mathematics was done without computers, until computers were ascribed cognitive functions – heuristic tools, tools for proving – in the production of mathematical knowledge. Astronomy was done without telescope, until telescopes were ascribed cognitive function in the production of astronomical knowledge. And often, the inclusion of new tools into distributed cognitive systems comes with new cognitive functions to ascribe to human agents; in our two examples, expertises in programming and in optical devices come to be required. Representations ascribing cognitive function determine the cognitive architecture of systems. They concern as much other elements of the systems as the one to which the function is ascribed, because they tell these other elements from where to take their input. It is the port bearing taker that provides the bearing timer with bearings. It is the telescope that provides enlarged images of celestial bodies, which the astronomer will take as input for further analysis. Ascriptions of cognitive functions determine information flow by regulating whom or what to trust for which type of data. The regulatory representations on which I will now focus are those that regulate epistemic deference in distributed cognitive systems.

10.4 How humans can distribute cognition

The cultural evolution of cognition is uniquely human, so there must be some cognitive mental ability that is proper to the human species and that is at work in changing the cultural conditions (or substance) of cognition. The biological evolution of human cognition is relevant for understanding the cultural evolution of cognition because its understanding may reveal what are these specific cognitive abilities that function as the motor of cultural evolution. Biological research on the evolution of cognition is very dynamic. The question evolutionary biologists ask is: why and how has the cognitive apparatus of different species evolved? How did the cognitive capacities contribute to increase the mean fitness of individuals in the species? What are the biological functions of the cognitive capacities?

It is important to distinguish the socio-cultural evolution of cognition and the biological evolution of the human cognitive apparatus. The latter obeys the principles of neo-darwinian evolutionism, while the former has principles proper to social and cultural evolution[2]. It is not as a source of an analogy for cultural evolution that I appeal to biological evolution, but because it has produced the human cognitive apparatus, without which cultural evolution would not be possible. Although the human genome is not very different from that of chimpanzees, human cognition is manifestly quite different from other animals' cognition. Two things that make human cognition so special are that it produces culture and that it is embedded in culture. This has lead Tomasello to talk about "the cultural origins of human cognition" (2001). In the book having this title, Tomasello traces what enables humans to develop cultures and what distinguishes their cognitive capacities from those of other species. The specifically human cognitive capacity, Tomasello says, is an adaptation for culture in the form of a *capacity to understand and share the intentions of others*. It is arguable that this capacity is the only one that is proper to humans, and it is also difficult to put a strict boundary between some apes' social cognitive

[2] Some theorists have tended to equate the principles of biological evolution and the principles of cultural evolutions. In particular, Dawkins (1976) has described cultural evolution as a Darwinian selection of cultural items, reproduced through imitation. The idea is also present as regard to the history of science in Cambpell's evolutionary epistemology (Campbell, 1974), where science is said to evolve through blind variation and selective retention of scientific ideas. The fact that cognition and culture co-evolve with social organisation shows that the application of Darwinian models for the evolution of culture and science is far from straightforward. As noteworthy, some developments in evolutionary biology are concerned 'niche construction' – the fact that species sometimes strongly contribute to the construction of the environment in which they evolve. Human construction of the socio-cultural environment can be associated with niche-construction (see Odling-Smee et al., 2003, esp. chap. 6). Sperber however, has developed a convincing argument against Darwinian models of culture: cultural evolution deeply and pervasively involves the constructive cognitive processes of the human mind, which transform rather than simply reproduce the representations that make culture. The analogy between neo-Darwinian biological evolution and cultural evolution is strongly limited because there is no replicator in cultural evolution as genes in biological evolution (Sperber, 1996a, chap. 5).

abilities and ours.

In any case, this 'adaptation for culture' is well at work in distributed cognition. Understanding and sharing the intentions of others enable people to take part in projects, and understand their own part of the work as distinct from what the partners will do. For instance, the preservation of distributed cognitive systems across generations requires that people learn from their predecessors how to take a role, a function, in distributed cognitive systems. The capacities Tomasello describes lead to "powerful forms of cultural learning, especially imitative learning in which the observer must perform a means-ends analysis of the actor's behaviour and say in effect 'when I have the same goal I can use the same means (action plan)' " (Tomasello et al., 2005). Understanding goal oriented actions is understanding actions as being functional. Most learning, in humans, is not mere reproduction of behaviour, it is the reproduction of means understood as means for a given goal; it is learning to perform functional behaviour. Human agents in distributed cognitive systems do have representations of the function they perform as goal oriented action plans. Even in the most uninformed situations, as when working on the assembly line in a Taylor type factory, people perform their task by knowing the means and goals – such as 'screwing a screw (goal) by doing this and that (means)', as opposed to a set of instructions without goals such as 'holding a screwdriver this way and turning this way'. Furthermore, Tomasello argues that humans have the ability to represent shared goals, joint intentions and the distribution of tasks for achieving the goal. This means that humans are able to engage in collaborative tasks and decide in situation the distribution of labour. For instance, two people carrying a large table through a door choose their positions so as to share the weight and carry it most effectively; they adapt their pace to each other; each of their move is dependant on the other's moves, so that the table does not to bump in the frame of the door. There is also, on top of the distribution of physical labour, a distribution of cognitive labour. When carrying a table, the carriers bear some responsibility for not bumping the side they carry, and they bear some responsibility for avoiding putting their partner in a clumsy situation. This means that they are given the cognitive task of representing their own situation, the situation of their partner, and the set of possible collective actions. Moreover, information flows between the carriers. It is not even necessary that the partners verbally communicate for information to flow. Both partners' cognitive processes are situated (their decisions always depend on the current situation rather than blindly carrying a pre-established rational action plan), and both partners frame the situation through their own action. Thus, they directly influence the situation of each other and constrain their respective set of possibilities. For instance, one carrier may rotate the table in a certain way, thus inducing a movement in her partner and putting him in front of new possibilities for pushing and pulling. A carrier moving the table communicates by the same token her intention as to the course of action that is to be taken. The distribution of cognitive labour is rendered possible by the fact that intentions and goals of others are understood, and actions are willingly adapted to situations that include others' intentions.

Collaborative actions are achievement that show the human potential to distribute labour, including cognitive labour. These achievements are creative and need to be distinguished from the highly collaborative work of social insects. For ants, collaboration is emerging from pre-established patterns of actions, activated in function of the actions of others. Labour is pre-distributed through the genetic determination of behaviour. In human collaboration, the way to collaborate is decided and negotiated by taking into account the intentions of others, deciding on a collective goal and distributing the tasks. Genetic determination bears on the cognitive capacity of understanding intentions and sharing goals, and leaves open the specific implementation of collaborative action. As a consequence, collaborative people can take up new goals and use new means adapted to the situations. They can distribute labour and cognition in many ways.

As we know, humans can also distribute cognition among non-human elements. The human ability to use things as tools is also species-specific, at least in its scope and achievement. I suggest that the human ability to ascribe tasks to artefacts relies on the abilities described by Tomasello, implementing similar cognitive processes to those involved in the distribution of labour to conspecifics. Using artefacts to achieve goals implies recruiting these artefacts in much the same way as one can recruit human agents. The artefacts are thought of as collaborators to the extent that they contribute to performing the tasks. It is because things are thought as possibly contributing to achieving a goal that things can be conceived as tools. Thinking of tools as meant to have a purpose is attributing them a goal, which is then shared by the users. The uniquely human creative way of using (and designing) tools would then be grounded in their abilities to share goals and to think of others as having intentions. Kurz-Milcke et al. (2004) description of 'cognitive partnership' would take a psychological signification if indeed humans think of cognitive tools with the abilities designed for thinking about human collaborators. In these situations, the ability to attribute intentions to others is explicitly used when interacting with cognitive tools.

At the basis of human distributed cognition is the ability to distribute cognition (this is not the case for social insects, where the distribution is determined by simple behavioural rules and the situation rather than by thoughts about how to distribute cognition). This ability involves deciding on the proper distribution of labour, and actions for the actual distribution. Such actions can be explicit communication, implicit communication or they can force the task upon others. These three types of action are located on a continuum and based on the idea that the goal is shared with the potential collaborators. This holds even when the more general goals are different, as when the employee and the employer want the same task to be done, but for different reasons. Communication is an action intended to generate the collaborators' understanding of the communicator's task ascription for achieving the shared goal. It is hoped that the collaborator will comply with the plan, but the distribution of labour can also be negotiated. Forcing the task upon a collaborator consist in putting him in a situation where the achievement of the shared goal requires him to take on specific tasks. The example of two people carrying a table is an instance where the limits between communicating and forcing are unclear, since the situation created by the carriers can be taken either as compelling or as suggesting courses of actions. In the case of cognitive tool use, the physical properties of the tool are understood, metaphorically but consistently, as intentions. And the tools are forced to comply with the goals of the users, because the users put them in a physical situation that 'makes them do' what the users want.

Human cognition, as any animal cognition, relies heavily on the environment: it is a situated cognition. On top of this, humans have the cognitive ability to decide which part of the environment they shall use for which purpose. Human distributed cognition is based on a mental process out of which cognitive functions are *ascribed* to elements of the environment, human agents or non-human agents. In most cases of situated cognition, environmental cues take a role in cognition without being represented as aid to cognition. Humans, however, are able to think of elements of the environment as aids for achieving cognitive goals. When using these aids, people distribute cognition. The goal of this sub-section was to begin the description of the psychological resources used for distributing cognition. Admittedly, I have only hinted at a research field, which is itself still exploratory. But it shows how deep into psychological research science studies could dig, in order to answer its most fundamental questions; and it points that the relevant psychological research is not bounded to the psychology of abstract reasoning. In the next sub-section I pursue this tentative description. It has been so far directed at the cognitive abilities that distribute cognition; I now focus on the representations that enable the distribution of cognition.

10.5 THE ROLE OF TRUST IN ASCRIBING COGNITIVE FUNCTIONS

In order to create a distributed cognitive system, one needs to distribute cognition not just for a single event, but for the achievement of an open-ended cognitive task. Cognitive systems such as brains, for instance, are meant to deal with the environment and determine behaviour *as long as* the organisms lives. And cognitive distributed systems such as the navigation system aboard a ship or scientific laboratories are meant to continuously produce cognitive output. Cognitive distributed systems are characterised by recurring collaborative actions that solve recurring cognitive problems. When distributed cognitive systems are created or changed, cognitive functions are ascribed. The ascription of a cognitive function implies the specification of a cognitive task and how to perform the task, and a specification of a place in a distributed cognitive system, including the formats of the input and of the output. Most importantly, the introduction of a new element in a distributed cognitive system requires appropriate change in the system so as to adequately integrate the new element. In fact, the new element will be functional only when it has been ascribed its proper role in achieving the general goal of the system. For instance, the microscope comes with a set of instructions about how to use it and for what purposes. It is used for providing pictures of microscopic living phenomena, which in turn are further processed (by a biologist's brain, for instance) so as to beget biological knowledge. Likewise, when a new employment is created in a lab, one provides the new employee with duties and means for his integration in the lab. This may include a job description and the required facilities, such as an office in the building of the lab. Introduction of theoretical and conceptual tools, such as notations, are no exceptions: they constitute elements that are used by the system thanks to a set of prescriptions – methods or rules – about how to use these notations. Within a distributed cognition analysis, these are all cases of creation of some new cognitive functions that are integrated within the previous cognitive systems. Changes in distributed cognitive systems can be analysed in two steps: first the advent of some new cognitive element, then the ascription of a cognitive function to this element, thus allowing its integration to the system. The ascription of cognitive functions passes through regulatory representations, as instantiated by the set of instructions often coming with a new cognitive tool. Not surprisingly, the function allocated is dependant on the environment constituted by the distributed cognitive system (more generally, the design of functions is always dependant on the environment). But the environment of the new functional element, i.e. the distributed cognitive system in which it is included, can be changed and adapted for the integration of the new element. Instructions about how to use a microscope, for instance, are changes in the practices of the scientists so as to genuinely integrate the cognitive tool.

When, in science, the ascription of the cognitive function requires drastic changes in the old system, this may cause a scientific revolution (in Khun's sense). This can happen because the object introduced requires a wide interface that allows the new object to have a pervasive action in the system. The introduction of Cartesian coordinates is of this kind. In the next chapter, I will analyse the impact of the introduction of computers in mathematics as implying changes in the social system for the production of mathematical knowledge: are computers just plugged in the system without requiring major changes in the traditional way to do Mathematics, or do they revolutionise mathematical practices? Note also that, according to this view, the advent of new conceptual frameworks and theories implies change in distributed cognitive systems. There are new sets of terms and symbols associated with rules, schema or models that prescribe the ways the terms can be used. Conceptual change in science can therefore be analysed as giving cognitive functions to terms, symbols and their associated processes in order to solve specific problems. The history of science, as argued in the previous section, has known numerous transformations of the cognitive systems that produce scientific knowledge. Artefacts, external representations, but also concepts and ideas can initiate transformations in distributed cognitive systems. Indeed, elements of cognitive systems form a continuous axis from theories to artefacts, where software or hardware aspects of the transformation are more or less prominent[3].

[3] When we translate these events to the distributed computation model, what happens is that (1) components are changed or added to the systems, and (2) the system is adapted by adding I/O instructions allowing communication with the components. The most obvious case is when one adds a printer to a computer. First one plugs the printer in, and then one provides instructions so that the previous computational system integrates the printer. To pursue the analogy further, the cognitive elements are pieces of hardware while the cognitive functions are the pieces of software that allow the hardware to effectuate its tasks. In some cases, the hardware is given a greater importance, such as when

In any case, all elements can integrate a cognitive distributed system and, henceforth, transform it as soon as they are ascribed a cognitive function. It is this very act of ascribing a cognitive function that transforms cognitive systems.

What is it that determines which elements with which functions are parts of distributed cognitive systems? The claim I now want to defend is that *trust is the 'cement' of distributed cognitive systems*. Trust is what holds the systems together, while changes in representations of what is trustworthy generate changes in the division of cognitive labour.

Trusting is what one does when accepting as true the information acquired by other means than one's own. Trust therefore refers to means of belief formation; it is both a psychological and an epistemic notion. The study of trust has generated an important literature in social epistemology (Hardwig, 1985; Coady, 1995; Goldman, 1999; Origgi, 2004, among others) and in the sociology of science (Shapin, 1995). The role of trust in scientific research has been shown to be pervasive and essential to scientific practice. In Relman's words:

[T]he fact is that without trust the research enterprise could not function... Research is a collegial activity that requires its practitioners to trust the integrity of their colleagues. (Arnold S. Relman, quoted by Hardwig, 1991)

Analyses of evolving distributed cognitive system could make great use of, and contribute to, this literature, which, in fact, addresses questions about what determines information flow in science. Indeed, trusting amounts to processing the information presented in the input by the agent trusted, and using this information for the production of the output result. Not trusting, by contrast, implies discarding the input provided by the

---

artefacts are introduced. In some other cases the software is more important, such as when new symbols and new theoretical terms are introduced. For instance, the meaning of some terms in physics that set the rules under which to use the term properly appear more important than the contingent term bearing the meaning. At the other extreme, the material details of some artefacts as simple as rulers appear essential to their functioning, but these artefacts always need some associated instructions.

agent not trusted. The consequence is that trusted elements are those elements that participate to distributed cognitive labour, while elements not trusted are not included in the cognitive processing out of which the cognitive goal is achieved. For instance, Hutchins reports the existence of a tool which is not trusted. It is the Omega. The tool has a computational procedure, it has an *intended* function: it measures the phase difference between the arrivals of signals from multiple stations. Omega was intended to provide accurate worldwide position-fixing capability. Yet, the Watch Standing Manual warns:

Caution: Positions obtained from Omega are highly suspect, unless substantiated by information from another source. In recent years, a number of costly and embarrassing groundings have been directly attributable to trusting Omega. No drastic decisions are ever to be made on unsubstantiated Omega fixes without the explicit permission of the navigator.

And Hutchins confirms that Omega is used only on rare occasions. He explains that Omega is a system that went into service before all the bugs could be worked out, and that it has been overtaken by other superior technologies. Whatever the reasons why the Omega is deemed not to be trustworthy, it is those representations about the trustworthiness of the tool that determine whether the tool will be included in the distributed cognitive system or not. With Omega, the answer is clearly no, in spite of an unsuccessful attempt to attribute it a cognitive function.

Representations of trustworthiness have two variables: the thing or person to be trusted, but also the kind of things it is to be trusted for. Hutchins gives us with the astrolabe an example of the relative independence of these two variables:

Sometimes, as the nature of the practice has changed, the role of particular instruments has changed. For example, the astrolabe was originally used both to measure the altitudes of celestial bodies and to predict the altitude and azimuth of a star. The observation-making duties were subsequently taken over by the quandrant [...]. The astrolabe [now] survives as the modern star finder [...]. It is used to get the setting of the precision instrument into the right neighborhood. It has been moved to a new job.

(Hutchins, 1995, p. 113)

Trusting something or somebody, therefore, is trusting it or him/her *for* something. Returning to the terms used up to know, ascribing a cognitive function is an act that involves both the thing or person to which the function is ascribed, but also the specific cognitive task that the thing or person shall take on. This point is most important for the case study we develop in the next chapter – the introduction of computers in mathematical practices: what are computers to be trusted for? Are they to be trusted for heuristic purposes only, for basic arithmetic calculations only, or also as proof makers?

People ascribe a cognitive function to some element (thing or person) when they systematically trust this element for solving specified cognitive problems. Consider the navigation team: the pelorus operator systematically trusts her alidade for the bearings it gives after being pointed at a landmark; the recorder trust the pelorus operator for giving the bearings of the landmark that was communicated to him; the plotter trust her hoey for reporting bearings on her map, and the bearing record log for providing the appropriate bearing; etc. The combination of these trusting behaviours enables the cognitive processes to span across tools and people, and their recurrent aspect makes the distributed cognitive systems.

It would be wrong to think that people working in a cognitive system have evaluated the trustworthiness of the element they trust. Trusting behaviour is directly determined by one's position within a distributed cognitive system. This is apparent for the navigation team, but it is also true in science. For instance, while the reviewer's work indeed requires doubting of some of the assertions of the submitting authors, the readers who are not expert in the field need not doubt the details of the argument; other expert readers may go as far as to attempt to reproduce the experiments. These differences are not (only) due to the rational nature of such trusting behaviour (whatever the norms of rationality that apply here), rather, the differences stem from the accepted distribution of cognitive labour.

But the question of trustworthiness is raised from time to time, and especially in problematic situations, as when the cognitive goal is not attained. At these points, trusting behaviour may be maintained after an evaluation of the trustworthiness of some elements, or trust may be withdrawn, and it may be given to some newcomer. When there is some change in some distributed cognitive system, then it can be assumed that some representations of trustworthiness are involved in the causal processes that govern the change. So, the question of trust arises more when framing distributed cognitive systems than when actually trusting in a particular occurrence. The history of evolving distributed cognitive systems is thus made of individual acts of trusting and occasional but decisive representations of trustworthiness. The question of trustworthiness arises not only for people, but also for cognitive tools. The problem consists in analysing the properties of the element (artefact or other) whose cognitive output is to play, or not, a role in the achievement of some cognitive goal: on the basis of these properties – competencies, honesty, and motives for human, physical properties for cognitive tools – one must decide whether to ascribe a cognitive function and which one. The ascription of a cognitive function to something within a distributed cognitive system involves more complex thinking than the attribution of a task for a single event. In the latter case, it is sufficient to foresee the behaviour of the potential collaborator at a given time in a given situation. The ascription of a cognitive function requires either controlling the long term situation, so that like situation engender like behaviour, or a better understanding of the element to which the function is ascribed, so as to have the guaranty that variations in situations will still trigger appropriate behaviour.

Regulatory representations for the *maintenance* of distributed cognitive systems essentially include representations of trustworthiness. These representations enable the information to flow from one element to another; they determine the architecture of the system by specifying where elements must take their input, and where they must deliver their output. Representations of trustworthiness have a decisive causal role when the distribution of cognitive labour, and thus the social organisation, is questioned. They have a key role when new elements enter a distributed cognitive system. A sailor taking a position in the navigation team will learn where to take her information from, i.e. whom to trust and for which data. Kurz-Milcke et al. (2004) show that this sort of learning is also essential to new members of their studied laboratory. They show that new members need to learn the past history of the modelling artefacts: what they have been used for, for which purpose they failed and in which conditions they succeeded. By learning the history of the artefacts, new members come to understand the conditions under which the modelling artefacts can be trusted. They elaborate mental representations of the trustworthiness of the artefacts, i.e. of their reliability regarding their cognitive output. Representations of trustworthiness and cognitive reliability also have a causal role in *changing* distributed cognitive systems. Recruitment of new members involves a thorough investigation of the trustworthiness of the applicants, as is manifested by their guaranty of expertise in relevant domains and of epistemic virtues. The design of new cognitive devices involves representations of the constraints that the device must meet. The problem of reliability essentially figures among these constraints. Last but not least: distributed cognitive systems have been defined as institutions endowed with a cognitive function. This endowment does not come from nowhere, it is the result of people *ascribing* a cognitive function to the social entity – and this is done through their representations of the reliability of the institutions. Such representations, as evaluating the value of the distributed cognitive systems, are importantly present in the feedback loops that maintain the systems in time. A tentative naturalistic characterisation of distributed cognitive systems follows:

A distributed cognitive system is an institution whose regulating representations include representations of trustworthiness of the institution itself, which take part in a causal chain constitutive of positive feedback loops.

Research in social epistemology has focused on the act of trusting testimony (Coady, 1995; Goldman, 1999). In some of these researches, the act of trusting is singled out and abstracted from the social context where it occurs, and it is divorced from research in psychology (e.g. Goldman, 1999, chap. 4). The result is a normative theory that assumes that scientists have some unrealistic super-competence for the assessment of trustworthiness. Avoiding such pitfalls is possible by being informed by work in cognitive psychology on trust (e.g Koenig and Harris, 2005; Clement et al., 2004), and by the re-description of the task to which agents are confronted in real life. The latter re-description is one of the goals of analyses of distributed cognition, leading to the study of what Hutchins calls "cognition in the wild". What cognitive processes determine trusting behaviour when in the wild? The relevant processes are distributed in time and across several cognitive elements: the trusting behaviour is determined by some representations of trustworthiness, but these representations need not be present in the head of the person who trusts. The person may trust because of his situation within a distributed cognitive system, while the system has been framed, and is currently maintained, by the relevant representations of trustworthiness. In these cases, the representations of trustworthiness are not proximal causes of trusting behaviour, but they remain key determinants.

10.6 CONCLUDING ON ANT

The objective of the last three chapters was to sketch a theory of the social and cognitive causes of organisational and technical change in the scientific labour process. The means of this theorisation are Sperber's epidemiology of representations and Hutchins' analysis of distributed cognition. When applying the latter to the study of science, I have used the work of Giere and Nersessian, which I pursued further with a sociological analysis of institutions endowed with cognitive functions, and a psychological analysis of the mental processes at work in the making of such institutions. The conclusions are as follow:

- Scientific knowledge is produced by distributed cognitive systems

- Distributed cognitive systems are institutions endowed with a cognitive function
- Institution making and maintenance is the result of processes which involve the causal role of regulatory representations and the emergence of macro-social properties.
- Among the regulatory representations, representations of trustworthiness or reliability are of major importance in the ascription of cognitive function
- Thus, representations of trustworthiness or reliability determine the maintenance or evolution of distributed cognitive system (through feedback loops), and the internal the architecture of distributed cognitive systems.

When focusing on the processes through which cognitive elements are integrated into distributed cognitive systems, I also follow the research trend and theories of Actor Network Theory (ANT) (e.g. Latour, 1993, 2005). ANT takes as main object study the constitution of networks in much the same way as I have focused on the constitution of distributed cognitive systems. Also, ANT makes "enrolment" the key events through which networks evolve. It has developed rich hypotheses about the processes of evolution of networks, as for instance the description of four stages where (1) actors or elements are identified, (2) they are enrolled (3) an identity to the new obtained network is created and (4) the network is mobilised. Giere and Moffatt's (2003) rendering of Latour's work in terms of distributed cognition is just a preview of what can possibly be done. But how useful would such a translation be? Assuming that the ANT has numerous interesting arguments and rich case-studies, what is the point of re-phrasing their texts? I have strongly opposed one much publicised methodological point of ANT, which consists in denying the distinction between human and non-human, thus closing the door to the import of psychology. This methodological point is unfortunately operating in a pervasive way in much of STS. Introducing terms such as 'ascription of cognitive function', 'mental representation of social institutions', 'representations of trustworthiness', etc. is opening the door to psychology. This is motivated by the increase in explanatory power obtained by putting at work a relevant body of empirical science – what happens in the mind has causal effect on scientists' behaviour, and thus on the making of science –, and this is motivated by a naturalistic programme, which does not sweep the causal power of people's mind under the carpet, but attempts to understand this causal power as stemming from natural (material) processes.

Theory and historical analysis, of course, go hand in hand. I have focused on theory so as to bring forth a synthetic account of the causes of changes in the organisation of scientific labour. In the next chapter, I confront the theory with a case study.

Chapter 11

Distributing mathematical cognition: the case of the 4-colour theorem

In 1852, Francis Guthrie, a graduate student at University College London, emitted the hypothesis that any map could be coloured with 4 colours only, and such that no country share a border with an identically coloured country. This conjecture resisted proof for more than a century and was eventually demonstrated in 1976 by Appel and Haken. Yet, the 4-colour problem does not end here for the proof of what is now known as the 4-colour theorem (4CT) essentially relies on untraditional means: computer analysis. The enormous amount of cases that need to be taken into account in the proof of the 4CT cannot be considered by a human brain. So the great innovation of Appel and Haken, and Heesch before them, was to relegate this task to a high-speed computer. Yet, this controversial move in mathematical practice has raised a great amount of debate regarding the status of the 4CT. Philosophers as well as mathematicians have wondered whether, given its unorthodox proof, the 4CT was really a theorem at all[1]. This led the philosopher Thomas Tymoczko to state, in 1979, the existence of a new four-colour problem, which concern the use of computers in Mathematics.

In this chapter I attempt to understand the new 4-colour problem by

[1] E.g. Bonsall denies it is a theorem and call for 'a proper proof' of the 4-color problem. C.f. Bonsall, F.F. 'A down-To-Earth View of Mathematics', *American Mathematical Monthly*, vol. 89 (1982), 8-15.

analysing it as an expression of change in the organisation of the production of mathematical knowledge. I use the theoretical apparatus of distributed cognition in order to further the analysis, and as methodological tool for the integration of social and cognitive studies of science. I also use this case study in the history of Mathematics to illustrate the points made in the previous chapters, especially regarding the principles of evolution of distributed cognitive systems. My hope is that the interpretation of the historical events in terms of evolution of distributed cognitive system will bring further understanding of the history of the 4CT. My historical account is based on second sources in the history of mathematics: I do not bring up new events, but only new interpretations of these events, questioning its historical relevance for the evolution of mathematical practices. Eventually, the strength of my arguments in the methodology of science studies partly rests on whether the theoretical tools brought forward can help make a relevant contribution to the history of mathematics.

In the first part of this chapter, I introduce the history of the 4-colour problem. A brief account of the proof itself is meant to give an idea of the task that was eventually given to computers, and at how mathematicians came to think that a possible means of checking the truth of the 4-colour conjecture was using to use non-human computing power. In the next section, I provide elements of an explanation of why so much effort was devoted to proving a rather useless conjecture. In the last section, I argue that mathematical cognitive practices are distributed, and that mathematics, as a discipline, constitutes a distributed cognitive system. I then apply this view to understanding the history of the proof of the 4CT: its historical significance is that it expresses an evolution of the distributed cognitive system that produces mathematical knowledge.

11.1 THE SIGNIFICANCE OF THE 4-COLOUR THEOREM

11.1.1 BRIEF HISTORY OF THE 4-COLOUR THEOREM

A brief account of the 4CT's long history is as follow (this account is drawn from second sources in the history of mathematics, viz. MacKenzie 1999;

Wilson 2002; Fritsch 1998):

It started with Francis Guthrie's conjecture, in 1852, about the colouring of maps with four colours only: it was hypothesised that by carefully choosing which colour to use for colouring each country, one could obtain a map where no country with a common border would have the same colour. The conjecture was communicated, via Guthrie's brother, to the University College professor of mathematics Augustus De Morgan. De Morgan was immediately interested by the problem, but the conjecture did not arouse much interest until 1878, when it communicated at the London Mathematical Society by Arthur Cayley's queries. One year later, Kempe published the first alleged proof.

Let us say that a map is 4-colourable when it satisfies the conditions of the conjecture, i.e. when it can be coloured with no more than four colours, without having neighbouring countries of the same colour (it must also be specified that countries must be connex and that countries meeting at a point are not considered as neighbours). Kempe's proof wanted to show that the property of 4-colourability was preserved when adding a country to any 4-colourable map. Unfortunately, his method for colour re-ascription, after adding a new country to a map, was shown not too preserve the requirement that two countries with a common border be of different colours. The counter-example was published by Headwood in 1890 - 11 years later. A pseudo-proof by Tait has known a similar fate: published in 1880, it was shown to contain a gap in the argument by Peterson in 1891. Both attempts, however, informed further research on the conjecture, and Appel and Haken's proof is using partial results from this earlier work (Kempe chains, 3-edge-colouring). Appel and Haken's proof is reformulating and improving the overall strategy, rather than taking a completely different path. In particular, the method of proof is still induction on the number of countries of maps.

It is worth taking a glance at the basic structure of the proof, so as to pin down where the problem arose, which founds its solution in computer work only. As said, Kempe's strategy consists in a proof by induction: assume that all maps with n countries are 4-colourable, then prove that a map with $n + 1$ countries is still 4-colourable. The initial step of the induction obtains with maps with less than 4 countries. The key idea of the proof consists in picking up a country, in a map with $n + 1$ countries, which has less than 5 neighbours. Re-colouring the map so as to find a colour shall be much easier for countries with few neighbours than for countries with many neighbours. For instance, if the map with $n+1$ countries has a country with less than three neighbours, then the inductive step is obvious: take this country out, the map obtained has n countries and is 4-colourable by induction, but putting back the new country preserves 4-colourability since it is possible to give to the added country the colour that is not used by its 3 neighbours. If the map with $n + 1$ country has no country with less than three neighbours but at least a country with less than four neighbours, then proving the inductive step similarly consist in taking out the country with less than 4 neighbours, using the inductive assumption that maps with n countries are 4-colourable, then finding out a colour for the country that has been taken out. Finding out this colour is not as straightforward as for the case of a country with three neighbours only because the four neighbouring countries may be coloured with 4 different colours. When that is the case, then a re-colouring of the map with n countries is needed, so that the four neighbouring countries use three colours only. This re-colouring is shown to be possible through a technique called Kempe-chain. The hard cases are the one where there is no country with less than 5 neighbours. It is upon these cases that mathematicians have worked upon during the 20$^{\text{th}}$ century, and it is for such cases that computer work is needed.

For the proof of the 4CT to go on as described above, one has to show that it is sufficient to prove the 4-colourability of maps that have at least one country with less than five neighbours. This is possible by calling on results in graph theory. The first step of the proof consists in showing that the 4-colour problem is equivalent to a dual problem formulated with graphs: take a point (call it a vertex) within each country and give it the colour of the country, trace a line (call it an edge) between points that belong to neighbouring countries. The 4-colour conjecture is then equivalent to the conjecture that vertices of graphs can be coloured with 4 colours only and such that no edge links two vertices of the same colour. For any graph, it is possible to obtain a 'triangulated graph' by adding edges between any two vertices, each time there can be such an edge that does not cross-over another edge. If a triangulated graph is four colourable, then taking edges out do not alter 4-colourability. So the four colour theorem can be deduced from the 4-colourability of triangulated graphs.

Dealing with graphs permits to apply Euler theorem (the number of vertices minus the number of edges plus the number of faces of a polyhedron equals two), and triangulated graphs have this further property that their number of edges and faces is a function of their number of vertices and the number of neighbours of each vertex. Out of these equations, one can conclude that any triangulated graph must contain vertices with five or less neighbours.

The inductive step for triangulated graphs that have no vertices with less than five neighbours require taking out of the graphs with $n+1$ vertices not a single vertex with less than five neighbours, but a whole sub-graph. The sub-graph together with the specification of how it is connected to its encompassing graph is called a configuration. A configuration is reducible if adding it to a graph preserves 4-colourability. So the challenge, now, is to find a reducible configuration in any triangulated graph. The cases where the triangulated graphs have a vertex with three or four neighbours have been dealt above: the reducible configurations are exactly the vertices with three or four neighbours. But for the case of triangulated graphs whose vertices with the smallest number of vertices is five, only larger configurations can be proved to be reducible. The strategy of the proof is to find a finite set of configurations – called an 'unavoidable set', such that any triangulated graph would include at least one of these configurations, then to work out the reducibility of the configurations of the set. But if the unavoidable set contains a configuration that cannot be shown to be reducible, then one must manage to find another unavoidable set that does not contain this configuration. In brief, the proof is done when one has found an unavoidable set of reducible configurations.

Techniques for showing reducibility of configurations were worked out by Birkhoff in 1913. It is these techniques that eventually lead to the use of computers. Heesch was the first to 'translate' the procedure for showing D-reducibility into an algorithm for computer implementation, and the computer programme was written by Karl Duerre in 1965 and implemented in Hanover on a CDC 1604A.

Much of the work in the proof, however, was dedicated to find an unavoidable set as small as possible. In 1948, Heesh estimated that an unavoidable set of reducible maps might have 10 000 members. Appel and Haken went throught a set of 1 476 maps, and Robinson et al. improved the proof in 1996, by finding out an unavoidable set of 633 maps. A key contribution is again due to Heesch, who designed a method for finding unavoidable sets, which is called the 'discharging procedure'. The procedure was also implemented by a computer so as to find unavoidable sets, but interestingly, Appel and Haken's proof do not rely on the computer. In other words, the action of finding the eventual unavoidable set of reducible configuration implies computer work. But once the set is found, proving that it is an unavoidable one can be done without computer work. One may say that the computer has an essential, but yet only heuristic role in this part of the proof. This use of computer has not been controversial because it does not need to be mentioned in the written proof. The procedures for showing reducibility of the configurations of the unavoidable set, on the other hand, rely on computer and the work of the computer needs to be included in the proof.

The problem readily perceived with the computerised approach was the time required for computation. A relatively small map of 13 countries at the outside skirt, would, at first, require between sixteen to sixty-one hours for checking D-reducibility; and more than a thousand of maps had to be checked. Duerre also had to punch the data onto punched cards. Much advance in the proof, at that stage, was due to the use of more powerful computers. Heesch and Duerre went to Brookhaven Laboratory (USA), thanks to the invitation of Shimamoto, chair of the applied mathematics department. At Brookhaven, Heesch and Duerre exchanged their CDC 1604A, programmed in Algol 60, with the Control Data 6600 designed by Seymour Cray and which was programmed in Fortran. The proof of Appel and Haken eventually was done using a IBM 370-168, programmed in assembly language.

11.1.2 The 4-colour theorem as food for thought for the students of mathematical knowledge production

Philosophers have focused on the significance of the 4CT for the nature of Mathematics and for defining the essential features of proofs. The most influential article is certainly Tymoczko's 'The Four Colour Problem and Its Philosophical Significance' (1979). In this article, Tymoczko argues that "if we accept the 4CT as a theorem, we are committed to changing the sense of 'theorem', or, more to the point, to changing the sense of the underlying concept of "proof"". What Tymoczko calls the new four-colour problem stems from the fact that nobody has *seen* a proof of the 4CT: mathematicians know that there is a proof of the 4CT only because a computer has told them so.

Tymoczko analyses three major characteristics of proofs: a) proofs are convincing, b) proofs are surveyable, and c) proofs are formalizable. While the characteristic of being convincing is recognised to be at the heart of the notion of proof, this characteristic stands itself in needs of explanation: why and how do proofs manage to be convincing? Being surveyable and formalisable are explanations of the property of being convincing in mathematics. Tymoczko argues that formalisability is an ideal criterion of surveyability: it analyses surveyability into finite reiteration of surveyable patterns. Surveyability has best characterised the traditional notion of proofs because *in practice* formal proofs where known only through the mediation of surveyable proofs: "Either the formal proofs are simple enough to be surveyed themselves and verified to be proofs, or their existence is established by means of informal surveyable arguments."(p.62). This means that even when one considers proofs as abstract formal entities, the actual proofs provided, those that mathematicians produce and assess, are first and foremost surveyable proofs. Thus, after having convincingly shown that surveyability is traditionally a central, if not essential, characteristic of proofs, Tymoczko observes that the proof of the 4CT is not, and cannot be, surveyable in its entirety. Appel and Haken's proof has *convinced* the community of mathematicians that the 4 colour conjecture is actually true. The proof provides reasons to believe that there exist a *formal proof* of the 4CT. But there is no *surveyable* proof of the 4CT and there is no surveyable proof that a formal proof of the 4CT exists, because the unavoidable work of the computer is itself not surveyable. The inclusion of the work of computers into proof has important philosophical consequences. Most notably, the use of computer-dependent proofs is ipso facto dependent on the empirical knowledge we have about hardware. The traditional divide between a priori mathematical knowledge and a posteriori empirical, scientific, knowledge is bypassed by the 4CT. The proof of the 4CT includes a lemma that is the result of the implementation of a programme on a computer, which is, in the last instance, an experiment. In Tymoczko's words:

> The appeal to computers, in the case of the 4CT involves two claims: (1) that every configuration in U [the set of unavoidable configuration,] is reducible if a machine with such and such characteristics when programmed in such and such a way produces an affirmative result for each configuration, and (2) that such a machine so programmed did produce affirmative results for each configuration. The second claim is a report of a particular experiment. (p.73)

At the end of his article, Tymoczko controversially concludes that the 4CT amounts to a change of paradigm in mathematics.

MacKenzie (1999), as a historian and sociologist of science, takes the 4CT to show that the concept of proof and thus the boundary of what constitutes mathematical knowledge are 'negotiable'. MacKenzie lists the followings reactions to Appel and Haken's 1976 article:

- Using computers in proof is just as using any other tool such as pen and paper. The proof of the 4CT is plain mathematics (e.g. Swart).
- Appel and Haken's solution to the 4-colour conjecture makes an illegitimate use of computers, because it renders the proof opaque to scrutiny. Appel and Haken's piece of work is not a proof of the 4CT (e.g. Bonsall).
- The 4CT has been accepted (de facto) as a contribution to mathematics. As a result the notion of proof has changed, since the proof of the 4CT introduces a new element in mathematical practice, viz. an empirical reliance on computer experiments (e.g. Tymoczko).

From the diversity of interpretations of the significance of Appel and Haken's use of computers, MacKenzie concludes:

> The Appel-Haken solution can be seen, therefore, as an anomaly, in Mary Douglas' sense. It was an entity that was hard to fit into the 'boxes' of accepted ways of thinking. Like all anomalies, the Appel-Haken solution raised the question of boundaries – in this case, the boundary between mathematics and the empirical sciences. (p. 8.)

In the tradition of the sociology of scientific knowledge, MacKenzie takes the controversy as revealing the social conventions that underpin what is usually taken for granted. The discussion over whether one should incorporate Appel and Haken's work in the corpus of mathematical knowledge bring to light the under-determination of word application – here, the word 'proof'; a theme that is dear to the sociologists of the Strong Programme. MacKenzie's historical work is intended as enriching the empirical and historical knowledge on the practice of proving, along the line of Lakatos' analysis of the history of Euler's theorem (1976). Thus, MacKenzie wants to contribute to the sociohistory of mathematics, which he defines as follow:

> By socio-histories I mean historical accounts informed by the type of question – concerning, for example, disputes over the meaning of 'proof' – that are of interest from the point of view of the sociology of knowledge. [...] To apply the sociology of knowledge to mathematics is to investigate whether mathematical knowledge is shaped by the social circumstances within which it arises: circumstances that might range from the nature of interactions between mathematicians to features of the wider society within which they work. (pp. 9 and 10)

My aim is similar to, and can be seen as a continuation of, MacKenzie's article. I also largely draw on his history of the 4CT. But, I want to add to sociohistory yet another dimension, the cognitive dimension, which is of interest to the understanding of scientific cognition and its evolution. Tymoczko and MacKenzie's works on the 4CT analyse the practices that produce mathematical knowledge, from a philosophical and sociological point of view respectively. I will add and integrate the cognitive perspective.

11.1.3 Significance of the 4-colour theorem for the cognitive history of mathematics

In her article 'Opening the Black Box: Cognitive Science and History of Science' (1995), Nersessian called for the development of 'cognitive history of science'. Cognitive history of science investigates the "thinking practices through which scientists create, change, and communicate their representations of nature" and examines "the cognitive tools scientists employ and the artefacts they construct in theoretical and experimental thinking practices". An aspect of scientific practice that is at once shown to be relevant for cognitive history is the use of cognitive tools. Scientists make an extensive use of instruments in their experiments, so scientific thinking is also, and importantly, thinking with instruments. Moreover, these instruments are used so as to provide information about the world and participate in the transformation of representations that eventually leads to scientific knowledge. Because they participate to the processing of information, these instruments qualify as 'cognitive tools'.

MacKenzie points out that social studies of Mathematics are very few compared to social studies of empirical sciences such as physics. Regarding cognitive history of science, Mathematics has also been neglected: works in cognitive history of science mostly deal with empirical sciences, mainly physics and chemistry (e.g. Nersessian, 1984; Tweney, 1991; Gooding, 1990). Also, while the archaeology of mathematical practice and the study of ethnomathematics do give attention to artefacts that serve Mathematical practice, the study of modern mathematics accord them little attention. For instance, Kitcher's framework for the comprehensive study of mathematics (1984, chap. 7)mentions the following five components of mathematical practices: a language, a set of accepted statements, a set of accepted reasoning, a set of important questions and a set of philosophical or metamathematical views, but no explicit mention is made to the use of tools or artefacts. The use of cognitive artefacts does not figure either in Bendegem and Kerkhove's enriched framework (2004). Similarly, Lakoff and Nunnez's inquiries into the embodiment of mathematical cognition (2000), and their ensuing theory of embodied mathematics, consider only the embodiment of the mathematicians' mind and ignore the important cognitive functions of tools for mathematical cognition.

What is the significance of the 4CT for the cognitive history of mathematics? The most salient event is certainly the introduction of computers in mathematical practices, which is highlighted by the dramatic presence of a reliance on computer within a proof. Computers are the paradigmatic example of cognitive tools, since their processing of information is so pregnant and obvious, and the study of human-computer interaction is largely using and contributing to the situated and distributed cognition paradigm (Hollan et al., 2000). The notion of 'distributed cognition' is, I will argue, an appropriate theoretical notion for describing the way mathematics is done with computers. The theoretical assertions of the previous chapters will be shown to grasp the historical events that pave the introduction of computers in the practice of mathematics — the history of the 4CT in particular.

11.2 MATHEMATICAL CONJECTURES HAVE COGNITIVE APPEAL

Studies on distributed cognition have mainly questions how cognition was distributed. In the previous chapters, I raised the question about why it comes to be distributed the way it is, for performing a cognitive task. It appeared that distributed cognitive system co-evolves with the cognitive goals, which can change as interests and opportunities arise. More generally, distributing cognition is not only choosing means for solving a task, it is also choosing the task to solve. How do we choose to which problems our cognitive resources should be distributed? In particular, why do we accord attention and cognitive effort to solving mathematical problems? How do mathematicians decide to which problem they want to work? How and why do mathematicians distribute their cognitive resources?

I argue that the historical evolution of mathematical research on the 4-colour conjecture was partly determined by the relations the 4-colour conjecture holds with human cognitive abilities. This relation is what made the 4CT so appealing to a number of mathematicians. The importance of the 4-colour conjecture may appear paradoxical with regard to its significance as a Mathematical theorem. Indeed, the 4-colour theorem is not a fundamental theorem upon which important mathematical developments would depend. It is not either the achievement of a surprising theoretical construction bringing to light deep mathematical facts. Moreover, the 4-colourability property is of no empirical or practical consequence[2]. Map makers have never wondered about how many colour where needed for colouring a map without having neighbouring countries coloured with the same colour. They just had sufficiently many colours and no reason to use 4 colours only. This insignificance of the 4 colour conjecture leads us to search for other reasons why the conjecture has been, and still is, so famous. For it is famous indeed. The conjecture is not only well known

[2] There are, however, analogies between the process of checking the design of computer chip and the 4CT proof.

among mathematicians of the 20th century, it is also among the most popular mathematical results. Moreover, the conjecture has had many mathematicians actually working hard on it, even spending their lives searching for a proof. It has had a reputation of 'man eating problem' and Appel and Haken estimated that ten millions person-hours had been devoted to the search of a proof (MacKenzie, 1999, p. 9, quoting Haken in an interview, and Appel and Haken, 1986). Eventually, the great popularity of the 4-colour theorem was boosted by the wide coverage of Appel and Hanken's work in newspapers.

Another peculiarity of the history of the 4CT is that many contributors to its proof are not usual classical mathematicians. The most striking of such figures is the French professor of literature, Jean Mayer, whose expertise in, and contribution to, the problem were much acknowledged. He was thus chosen by Haken as a reviewer of Appel and Haken's first publication of their proof. More central figures still do not fit in the archetypical image of the great mathematician. Kempe, although trained in Mathematics at Cambridge, pursued a legal career. Shimamoto and Swart, who worked on the computerised solution of the 4CT, were respectively a theoretical physicist and a physical chemist. Heesch and Haken, the two main mathematicians who favoured the exploration of a computerised proof, were mathematicians with important results: Heesch solved the tiling problem (Hilbert's 18th problem) in 1976 and Haken the Knot problem in 1954. But none had straight successful careers: Heesch first met difficulties in his early career because of his lack of sympathy for the Nazi regime, and then never attained the rank of full professor. And Haken characterised himself as being "really not a mathematician: I could not pass any one of those exams which are now required". His education in Mathematics was disturbed by the War (mathematic books were not available) and he worked as an engineer, after his Ph.D, for more than 10 years before being enrolled at the University of Illinois. This points to an important characteristic of the proof of the 4CT, which is that it does not involve highly technical mathematics and complex concepts (relatively to other famous conjectures).

How can we explain this particular history of the 4-colour conjecture? A factor that has probably partially caused the apparently unjustified fame of the conjecture is to be found in the facility with which one can understand the conjecture. There is no appeal to technical notions of mathematics in the statement of the conjecture, which consist in a relatively short sentence with little surprising information. Arrangements of colours on two dimensional surfaces, and the constraint on the colouring of neighbouring countries, are things that can be easily imagined. This provides easy content and straightforward illustration of the statement of the conjecture. It is hard to find a theorem whose content is as easily intuited. Yet, its surprising aspect calls for attention. Why four rather than some other number? Can't we imagine some map which is complex enough so as to provide a counter-example of the conjecture? From a cognitive point of view, the 4-colour conjecture has all the ingredients to make it popular: since it is easy to understand, it is easy to communicate; also, its surprising aspect catches attention. Boyer (2001) explains the cognitive appeal of religious ideas by the fact that they similarly combine common sense beliefs and surprising facts. His analysis of what makes an idea memorable does apply to the 4-colour conjecture. So while the 4-colour conjecture seemed to be rather insignificant in terms of applications and hindsight, seen from outside of mathematics, it is nonetheless a fascinating conjecture. The four-colour conjecture grasp the attention because little cognitive effort is needed to understand its stakes — topological representations of coloured map are easily imagined. On the other hand, the assertion is surprising because one contrasts easily very complex maps with numerous countries neighbouring each other and the little means provided to colour them. There are certainly numerous reasons why a mathematical proposition attracts attention. Another reason of the popularity of the 4-colour conjecture may lie with the combinatorial aspect of its proof. One can observe an apparently transcultural fascination for combinatorial problems: they are involved in many games, from the game of Go to crosswords. The incredible success of Sudoku is a case in point. The proof of the 4-colour theorem involves such combinatorial processes. The fact that combinatorial problems are catchy might be explained as a result of the illusion that the solution, i.e. the reward, seems always very accessible. The 'man eating' history of the 4-colour theorem is certainly not fully independent from the attractiveness that combinatorial problems have. According to the cognitive principle of relevance (Sperber and Wilson, 1986; Sperber, 2005), "the human cognitive system tends to allocate resources to the processing of available inputs according to their expected relevance." The relevance of an input increases as the processing of this input yield cognitive benefits — it issues new information; and the relevance of an input decrease as the processing of this input requires cognitive effort. In combinatorial problems, one always gets the sensation that the solution is at hand. It is just one step further and the whole thing will be clarified. Moreover, calculating the next step is the result of some algorithmic process — one just needs to check a few more cases. The consequence is that combinatorial problems provide inputs with high expected relevance. This hypothesis about the cognitive properties of the 4-colour conjecture and the human cognitive system, if true, would explain the historical success of the 4-colour conjecture. Sperber (1996a) proposes to study cultural phenomena in terms of the distribution in time and space of public and mental representations (entities with semantic contents which are respectively outside and inside the heads). He distinguishes two factors of distribution: the environmental factors of distribution of representations, such as the availability of writing as a means to communicate; the other factor of distribution is related to the cognitive processes of the human mind. Among other things, if a representation is attention-grabbing, easily remembered and communicated, then the probability that this representation shall be well distributed increases. In our case, the 4-colour conjecture does trigger cognitive processes in a way that favour diffusion: it is easily understood and remembered, and it conveys a high expectation of relevance. Of course, the cognitive aspects of the problem are not sufficient to explain its success: environmental factors such as the existence of a community of mathematicians sharing the problems that interest them is also essential. For instance, Cayley's 1878 query is known to have much contributed to raise the interest in the conjecture because it was addressed to a whole community of important mathematicians. Contingent events, such as Haken attending Heesch' seminar at the University of Kiel, where he was a student, and the ensuing contacts he has had with him, are processes of diffusion of the conjecture and elements of its proof that are part of the history of the 4-colour theorem.

With the introduction of computers in mathematical practices, a cognitive stake is the 'enrichment' and transformation of a cognitive practice with devices from information technology. Yet, the above cognitive analysis is still of significance for understanding the advent of computers in mathematical practice: the popularity of the conjecture was a reason why mathematicians working on it have been granted expensive computer time, and probably, the relative accessibility of the elements of the proof has facilitated strong collaboration with 'programmers'. The 'recruitment' of computers for the proof of the 4CT was facilitated by the accessibility of the problem. Further questions that relate, in our case study, the epidemiology of representations to distributed cognition are:

- What kinds of ideas about computers and their functions, and about mathematics and the nature of proof, led to giving computers a cognitive role in the action of proving?
- Why and how did these ideas spread?

I will attempt to answer these questions in the last section. But first, in the following section, I want to argue that the practice of mathematics does implicate distributed cognition. It will then be more easy to understand why and how computers have been attributed a cognitive function in the making of mathematical knowledge.

11.3 MATHEMATICAL COGNITION AS DISTRIBUTED COGNITION

This section is an application of the claims of the previous chapters to the discipline of mathematics. I argue that the practice and the history of mathematics can be better understood when the discipline of mathematics is described as a distributed cognitive system. Mathematical practices can then be seen as contributing to the overall function of the discipline of mathematics; and the evolution of mathematics can be fruitfully interpreted as the evolution of a distributed cognitive system. The analyses of social institutions and organisms as functional entities having cognitive goals inform the sociology of mathematics by specifying the nature of the social phenomena (functional, dealing with information) and thus the kind of sociological explanation required. The framework of distributed cognition thus can also benefit to cognitive psychology of mathematics — analyses of how cognition is distributed inform cognitive psychology by specifying the tasks that impart to the mathematicians' minds.

11.3.1 Mathematical cognition involves non-mental representations and social interaction

Is mathematical cognition distributed? Or, in other word, can we fruitfully apply the framework of distributed cognition to the analysis of the practices of Mathematics? The traditional image of the mathematician is that of a lonely thinker disconnected from the world. As opposed to the contemporary scientist who is equipped with a large battery of tools for scrutinizing and probing the world, the mathematician seems to need nothing but his own cognitive power to access the platonic world of the mathematical realm. And yet, mathematical cognition is essentially distributed.

Mathematical cognition outside the brains

One first fact that shows that mathematical cognition is, right from the start, making use of written symbols. As a matter of fact, mathematical cognition has provided a key early example that shows how some cognitive tasks are achieved by exploiting the environment. McClelland et al. (1986) point out that multiplying two three-digit numbers is rarely done in one's head. We use pen and paper. And when proceeding to the resolution of the task, we constantly rely on what we had written before. We learned to produce external representations, strings of figures, which need to be arranged in specific ways on the paper so that, for instance, units, tens, hundreds, etc. form columns. The computation is therefore done both within the head *and* through the action of writing numbers (maybe using one's fingers for the carry over). Thus, the most basic mathematical thinking relies on cognition that is distributed among mind's processes and the paper on which the operations are written. Proving also requires creating and manipulating external representations. It consists in constructing step by step a physical representation of a proof, and each partial representation of a proof is used for the cognitive production of the next step of the proof. The cognitive processes of proving therefore go back and forth between the mathematician's minds and external representations of partial proofs. All mathematical thinking includes the production of external representations that serve further thinking and further production.

When retracing the history of cognitive science, Hutchins (1995) denounces an original mistake that has been done by those who created the field: they took the model of the mathematician manipulating symbols as a model for what the mind is doing. The story goes back to the 'creation myth of cognitive science', which places the seminal insights of Alan Turing in his observation of his own actions, how he went about solving mathematical problems or performing computations. Hutchins comments that:

> Originally, the model cognitive system was a person actually doing the manipulation of the symbols with his or her hands and eyes. The mathematician or logician was visually and manually interacting with the material world. A person is interacting with the symbols and that interaction does something computational. This is a case of manual manipulation of symbols. [...] The properties of the human in interaction with the symbols produce some kind of computation. But that does not mean that that computation is happening inside the person's head. [...] What Turing modelled was the computational prop-

erties of a sociocultural system.

Hutchins' account of the history of cognitive science points out the original ignorance of the distribution of cognition that is done in mathematical practice. It denounces the attribution fallacy that follows: squeezing into the head of the mathematician what is in fact done not only with his brain, but also his hands and eyes, and pen and paper. The attribution fallacy is also worth denouncing because it is a source of psychologicism, the long denounced theory that asserts that the truths of Mathematics are psychological facts (see section 7.1). In the face of the initial attribution fallacy, Hutchins suggests that the modeling of the mind should not be in the form of computation-as-symbols-processing, which would only appropriately model some systems of distributed cognition. Hutchins favourite alternative, as other theorists of distributed and situated cognition (e.g. Andy Clark) seems to be some connexionist model of the mind (he mentions our pattern-matching abilities). However, connexionist models are not the only alternative to the model of the mind as a perfect mathematician processing external symbols: another possible theoretical choice is that of the mind as made of organised domain specific abilities (see chap. 5). In any case, the understanding of mathematical cognition is radically changed with the analysis of its distributed aspects; it has some consequence on the *psychology* of mathematics. The usage of external symbols for cognition is much more than an increase in memory space (the written space as a container of memories) or processing power. With the introduction of the new representational medium constituted by mathematical symbols, the cognitive processes are themselves qualitatively changed. For instance, the introductions of numerous mathematical symbols cannot be reduced to the mere encoding of pre-existing thoughts. New mathematical symbols come with their own syntactic constrains that determine cognitive processing. Think for instance of laws such as associativity, reflexivity, the multiplication of same numbers with different exponentials, the procedures to integrate and derive that come with the symbols $\int$ and dx/dy. Writing also provides mathematics with the property of being surveyable that goes far beyond the surveyability of spoken arguments: the scrutinizing can bear on the most hidden details of the reasoning, and focus attention on each step of the argument. With writing, the survey can be exercised on physical tokens (e.g. a written piece of paper). It is therefore not limited in time by the short-lived utterance; the survey can even continue across generations. Spoken discourse does not have the surveyabiliy that Tymoczko shows to be central to the traditional practice of Mathematics.

The importance of the distributed aspect of mathematical cognition is made apparent with the history of the production of external symbols (see Goody 1977). Indeed, the usage of external symbolic tokens to serve cognitive goals seems to have started with iconic clay tokens representing quantities. These symbols were used to serve mathematical operations for commerce, and preceded the cuneiform script of the late fourth millennium B.C. – the oldest known system of writing (Schmandt-Besserat, 1992). Mathematical symbols would therefore be the first artefacts used as tools for cognition, and mathematics would be the first practice that clearly and systematically distributed cognitive operations to non-human entities. Diagrams can also play an important role in Mathematical cognition. Zhang has shown that diagrams can play a direct role in cognitive processing and that they can change the cognitive processes at work in problem solving (e.g. Zhang, 2001, 1997).

Let us now turn to our object study: computers. In a trivial way, they are involved in mathematical cognition, since they now often replace pen and papers. But even this trivial fact may have some deeper consequences: hasn't LaTeX — the most popular software used for writing texts with mathematical symbol – somewhat changed the practice of doing mathematics? More obviously, computer generated graphics and drawings provide a powerful and widely used tool for mathematical cognition. In many cases, the representations output of computer processes are used as heuristic tools not figuring in the proofs. But the fact that they do not figure in the final output of the cognitive processes does not mean that they do not play an important role in cognition. Mathematics teachers are well aware of this, as their pupils manipulate their calculators to guide their mathematical reasoning. It appears that the heuristic use of computers and calculator is far from obvious and requires by itself specific knowledge and skills (Guin and Trouche, 1998). The place of computers in mathematical cognition is, at all levels, a delicate and important question — what is the specific role that computers can have in mathematical distributed cognition? Eventually, computers are used in theorem proving, such as in the 4CT. They are now used with increasing frequency in mathematical proof, with three better known proofs: the 4CT, the proof that there is no finite projective plan of order ten (published in 1889) and the solution of the Robbins Problem (published in 1997).

In brief, artefacts are used in Mathematics in order to store mathematical representations (as do notes on the paper), propagate them (as do scientific journals), and transform them (as do computers and other acted upon symbolic systems such as the abacus). Mathematical cognition essentially and pervasively relies on external artefacts.

The social distribution of mathematical cognition

I hope I have now shown that mathematics is a cognitive activity that relies on external artefacts. But what about the *social* distribution of cognition? In Hutchins' example, the cognitive task of navigating is not only implemented in the manipulation of cognitive artefacts, it is also socially distributed among human agents. And while science is well recognised to be a collective activity, the myth of the lonely mathematician may again prevent the recognition of the social distribution of cognitive labour in Mathematics. So I now want to argue that mathematics is a collective activity with socially distributed cognitive processes.

There is, to begin with, an increasing number of co-authored papers. Teamwork, Hardwig (1991) notices, is "not unknown in mathematics [. . .], due to the many areas of specialisation required to complete some proofs". Hardwig exemplifies this claim with Louis De Brange's proof of Bieberbach's conjecture. The proof, indeed, called for the conjoined expertise and work of De Brange, Gautschi, a 'numerical mathematicians' who checked inequalities involving hypergeometric functions and Askey, who was, at that time, the expert on special functions. It is also justified considering that the distribution of cognition among mathematicians spans through time, since cognition is always timely and no time scale enters the definition of distributed cognition. In that perspective, the resolution of a specified problem, such as the 4-colour conjecture, is a case of historically distributed cognition. The 4CT is the output of cognitive processes that span one century and which are distributed among numerous mathematicians: Kempe set the general strategy of the proof (1879), Birkhoff developed the techniques for proving reducibility (1913), Heesch developed the discharging procedure for finding unavoidable sets, and eventually Appel, Haken, Koch and the University of Illinois' IBM 370-168 provided the output proof. Hutchins has a similar point for the computations involved in position fixing: Given that the publication of the chart, the nature of the plotting tools, the mathematics of the projection of the chart, and even the organisation of the sexagesimal number system, "make as large a contribution to the computation as any other, we may wonder where we should bound the computation in time." Reviewing the possible bounds, Hutchins concludes that none are justified and that "we will not understand that computation [of plotting] until we follow its history back and see how structure has been accumulated over centuries in the organization of the material and ideational means in which the computation is actually implemented" (1995, p. 168). In Mathematics, nearly all computations make use of knowledge acquired and representations developed in the past by some other mathematicians. This is another reason why mathematical cognition is to be understood as distributed cognition. It should be noted in passing that the fact that cognition may span over very large period does not annul time constraints in cognition: position fixing should be done before the ship meets a rock. In mathematics, time constraints are less obvious but are still present. Time constraints are set by competition among mathematicians and by general expectation of results from society at large. In the case of the 4CT, Appel and Haken have been racing with other mathematicians working on the same project: Allaire, Heesh, Stromquist and Swart, who where probably no more than a year away (Mackenzie, 1997, p. 37).

At the organisational level of Mathematical practices, the 'unity' of the corpus of mathematical knowledge is such that their exists an important *interdependence* of the work of mathematicians. Some result in Algebra may prove to be essential to some proof in, say, functional analysis. So analysts may, and actually often do, use the work of other mathematicians. Mathematical cognition is therefore distributed among specialists in their fields. Each specialist has his own *cognitive task* — solving as many problems as possible in his branch — that often depends on the output of others specialists' cognitive processing — their theorems and theories.

The organisation of mathematical knowledge production, however, does not end with theoretical specialisation. It also involves the institutions that provide teaching in Mathematics, the apparatus with which mathematicians communicate and the complicate processes of evaluation of results and people. Cognitive functions are the one of students and teachers, conference organisers, editors and referees. Each function has its associated task and the way to proceed. Eventually, one can see a *cognitive architecture* that is being implemented in the institutions that enable mathematical knowledge production: representations of specific kinds are being processed in specified loci (e.g. the brains of specialists), they also follow predetermined routes of information (e.g. from a brain to some journal's printed article, passing through writing up processes, referees, editing). Conclusion: mathematical cognition implies processes distributed across mathematicians.

What about external structures? Do they play a role in mathematical cognition? A positive answer is already made apparent with the role of institutions within which mathematics is embodied. A department of Mathematics, for instance, is more than a set of pure minds. It includes the building itself within which moves the bodies of mathematicians, carrying their representations with them, which may be processed and communicated in offices, conference rooms, through e-mails, etc. Mathematical cognition takes place in and with the environment, which makes possible and at the same time constrains mathematical cognition. A department's production may be constrained by such contingent and material facts as whether the departmental library receives some specialised journal or contains one particular book. Eventually, it may even depends on the arrangement of the offices (it is easier to communicate with the colleague whose office is next door), the size of the conference room (the type of communication is not the same when the audience is small than when it is large), etc. All these facts should be well known to those who administrate research. The important funding problem is a good indicator of the embodiment of mathematical cognition.

Mathematical cognition is distributed both among mathematicians and among artefacts. The hardware, so to speak, of the cognitive processes that produce Mathematical knowledge is to be found both in Mathematicians' brain and in the external world. In Clark's metaphorical language, the mathematicians' minds leak and extend over artefacts, institutions and colleagues. Mathematical cognition is therefore distributed. In the next sub-section, I ask whether mathematics, as a discipline, form a distributed cognitive system. The notion of distributed cognitive system is stronger than the notion of distributed cognition: it includes supplementary assertions about the organisation of cognition and refers to social cognitive entities. It is to these aspects that I now turn.

11.3.2 THE CSM: THE COGNITIVE SYSTEM THAT PRODUCES MATHEMATICAL KNOWLEDGE

Academic disciplines as distributed cognitive systems

There exists several ways to characterise and identify scientific communities. Laboratories, institutes and research centres provide an obvious set of scientific communities. These research units are relatively small social entities with high degrees of communication and it has been shown that such social entities constitute distributed cognitive systems. But there are larger, encompassing, research communities that may also qualify as distributed cognitive systems. One interesting way to identify such larger scientific communities consists in analysing citations and co-authoring of journal articles. Interestingly, such analyses enable drawing clusters of highly connected authors, which furnishes a precise and timely image of fields of research (Small and Griffith, 1974). The results show not only clusters made of the classical academic disciplines, but also smaller and stronger clusters corresponding to specialities within a discipline or to interdisciplinary research topics. The interpretation of the clustering of citations and co-authoring is straightforward: people collaborate and cite each other when they have similar research interests, i.e. when their research bears on the same class of phenomena and uses the same means of investigation. Do such research fields qualify as distributed cognitive system? A positive answer is prompted by the fact that a research field has a cognitive goal (which is not necessarily well defined) and has large information flow between its elements. The distribution of cognition does not necessarily imply that different cognitive tasks are attributed to different elements. Different elements can contribute, in parallel and concurring between each other, to the same problem solving task, bringing their own specific means in the process. Knorr Cetina (1999) has identified, in science, different epistemic cultures, defined as "those amalgams of arrangements and mechanisms bonded through the affinity, necessity and historical co-incidence which, in a given field, make up how we know what we know. Epistemic cultures are cultures that create and warrant knowledge." (p. 1). The term of 'epistemic culture' is too broadly defined with regard to the notion of 'distributed cognitive system'. Giere (2002b), however, has shown that Knorr-Cetina's case studies – one laboratory in High Energy Physics and one laboratory in Molecular Biology – do qualify as distributed cognitive systems. Giere also eventually suggests that the two epistemic cultures identified by Knorr-Cetina correspond to two types of distributed cognitive systems. The framework of distributed cognition, indeed, provides interesting criteria for identifying scientific communities.

I suggest that academic disciplines are distributed cognitive systems. 'Academic discipline' is defined as a branch of knowledge and instruction, but the concept is often explained by its extension: academic disciplines are physics, chemistry, biology, mathematics, philosophy, etc. Each of these can refer to a body of knowledge, to the institutional settings of universities or research institutions, including the geographical repartition of the staff in buildings and the official curricula for students, to an abstract understanding of the methods and practices used for investigating a specified set of phenomena, etc. Dictionaries emphasise instruction or education, with the Latin origin of the world which meant instruction of disciples, but current debates in the management of science about 'interdisciplinarity' often talk about the institutional structures of scientific knowledge productions, which includes disciplinary journals or grant agencies (Origgi et al., 2004). This rich understanding of 'disciplines' may be interpreted as different focus on some of the aspects of distributed cognitive systems. There are the material aspects of academic disciplines, which are found in the dedicated buildings, libraries, and the set of teachers and practitioners of the discipline. The knowledge of disciplines is either the output of the distributed cognitive system or the set of conceptual and theoretical tools used in its computations. Teaching can either be understood as fulfilling a proper function of disciplines, producing people that are knowledgeable in the discipline, or as a sub-function of the system dealing with the reproduction of its own elements. Disciplines are social organisations. They do not have centralised headquarters but their internal actions or cognitive processes are still well organised for the production of disciplinary academic knowledge. These international decentralised social organisations are distributed cognitive systems

Taking academic disciplines as distributed cognitive systems also suits well our intuition that disciplines have some kind of autonomy and develop their own internal history. It thus provides a way to address the internalist-externalist debate in the history of science. Usually, cognitive systems are evaluated by external means: for instance, evolutionary psychology takes it that the maintenance of a mental cognitive device through generations depends on the added fitness it provides to the organisms. There is no representation of the fitness, but the there is nonetheless a feedback that directly depends on environmental factors (fitness is relative to an environment). When the function is a social function, by contrast, the standards are set socially. The goals of social distributed cognitive systems are set through social processes. For instance, contemporary navigation serves goals and complies with constraints that are specified with the needs of culturally situated people — e.g. people who are willing to sail for long distance and have the technical means to do so. As a consequence, the specific implementation of the European navigation system is dependant upon cultural and social history (as pointed out by Hutchins 1995, p. 115). Importantly, representations of social cognitive systems enter in the feedback process — the process involves evaluative representations of the performance of the system. The maintenance of academic disciplines also involves evaluative representation of the performances of these disciplines. But another particularity comes in: this evaluation is mainly internal, i.e. done by elements of the discipline itself. Self-evaluation is a specific and important feature of current scientific practice; and a particularity of academic disciplines is that feedbacks are mainly 'monitored' by the academics themselves. Gibbons et al. (1994) argue that the process of evaluation is more and more involving society at large. Scientists must now account of what they do, not only to their own colleagues, but also to research funding bodies (esp. industries or public institutions), to commissions for the ethic of science, lobbies, etc. The researcher is made socially responsible and his autonomy is lessened. The stakes are the boundaries of distributed cognitive systems and the processes through which these systems are maintained. These boundaries are social construct that evolve through time, depending on the interests and understanding of the people involved in the feedback loop. Distributed cognitive systems of science evolve, and so do disciplines. They have no perennial essence that guaranties their existence, and their autonomy is always relative. Qualifying academic disciplines as distributed cognitive systems implies recognising that scientific institutions are 'well designed'. But such recognition raises further questions: 'designed for what?', and 'how is the design achieved?' The answers cannot be bound to 'designed for the production of truth', and 'design achieved by rational decisions'. The sociological investigation of functional social systems requires investigating the actual social processes through which design is achieved — as produced by intentional action, but also from emergent properties, negotiated decisions, and through feedback processes.

The function of Mathematics (or why do we keep doing it)

Why and how mathematics, as a distributed cognitive system, is maintained through time? It must be because it perform a cognitive function (chap. 9). What is this function, and what goal does it achieve? This sends back to the classical questions about the relations between science and society: why do sponsors fund mathematicians? Why do politicians ascribe public funds to mathematical research? And to which extent do funding bodies determine the developments of mathematics? The history of the 4CT illustrates the complexity and historical contingency involved in the answers to these questions, which arise dramatically because the cost of computer usage was very high. Haken estimates that "at that time [early 70'], using a big computer was something like \$ 1,000 per hour"[3], and no computer was especially built for mathematicians. Why, indeed, attributing funds to provide computers to mathematicians? In the domain of computer science, Mathematicians were service providers rather than consumers. Some mathematicians, however, managed to draw on computer resources not primarily intended for them. The Control Data 6600 at Brookhaven, used by Heesch and Duerre in 1969 and 1969, was meant to be used for atomic physics, and the IBM 370-168 used by Appel and Haken at the University of Illinois was meant to manage administrative data. The attribution of computer resource was thus negotiated between mathematicians and other scientists. For instance, the Control Data 6600 was managed by Shimamoto, a theoretical physicist with an interest in the four colour conjecture. Heesch gained access to the computer resources via a social network: Haken attended one of Heesch's talks in 1948; Haken

[3] Quoted by MacKenzie (1999, p. 33), Interview by J. Dale in April 1994.

had a tenured at the University of Illinois and could talk to the head of the department of computer science at that same University, John Pasta; John Pasta talked to the chair of the applied mathematics department at Brookhaven, Shimamoto. Appel, as a mathematical logician, was pretty close to the computing community and, as he himself pointed out, "had friends throughout the computer establishment"[4]. In spite of the reluctance of the administration, he managed to get permission to use the University administration's computer during some of its spare time. This, Haken notices, was made possible thanks to Appel's "political skills"[5]. Why, then, did Mathematicians gain access to computer resources? There are the historical events and their particularities described above. Among the deeper reasons, however, one can designate the proximity of Mathematicians and computer scientists, both in terms of knowledge — which convinced computer scientists that computer resources would be fruitfully used – and in terms of relations – as shown in the above chains of relationships from Heesch to Shimamoto. The relations from society to the sources of funding need not be straightforward. Here the funding for mathematical research, as needing computer time, resulted from decisions of colleague scientists. Yet, the human time devoted to the problem has been permitted by the relative independence of the tenured position, which opened the possibility to devote large amount of time to the 'man eating problem' that was the 4-colour conjecture (e.g. Appel and Haken benefited from a tenured). However, it is also worth noting that numerous protagonists of the history of the 4-colour theorem were not professional mathematicians (Mayer, Kempe, Shimamoto, etc.). These people devoted their time and effort to the 4-colour conjecture even if it was not their social responsibility. This is the case of many mathematicians, especially before the advent of special institutions for scientists and mathematicians (Universities, research institutes, etc.). These facts point towards one single representation as the motor of the practice of mathematics, viz. 'mathematics is a fine and worthwhile enterprise'. Why Mathematics is seen so

[4] Quoted by MacKenzie (1999, p. 36), Interview by J. Dale in June 1994.
[5] Quoted by MacKenzie (1999, p. 36), Interview by J. Dale in April 1994.

is yet another difficult question, which calls for more empirical investigations — possibly illuminated by interest theory. I can just remind here that the usefulness and applicability of mathematical results was certainly *not* the motive that drove people to participate, fund, and sustain the search for a resolution of the 4-colour conjecture.

That 'mathematics is a fine and worthwhile enterprise' is certainly the main belief that drives the positive feedback loop that maintains the CSM. Interestingly, this belief has survived through many different historical contexts, drawing on different means of justification, from the ideal that mathematics is the language of nature, to the belief that good reasoning abilities could be acquired through the practice of mathematics. An important factor that sustains the positive reputation of mathematics is the cognitive pleasure that practitioners take out of it: there is a psychological factor of attraction towards the practice of mathematics. This said, another sociological question comes in: Who is to decide – and how – *what* is Mathematics? Two points are present in this question: first the normative character of mathematical knowledge production: not everything can join the corpus of mathematical knowledge. Mathematicians can make mistake as when Kempe and Tait published in 1879 and 1880 respectively, their erroneous proofs of the 4-colour conjecture. The second point raised by the question concerns the social processes through which an item enters the corpus of mathematical knowledge: these processes are distributed, and involve achieving a relative agreement among the community of mathematicians. The normative character of scientific knowledge, as present in the pervasive evaluation of mathematical practices and results, forms a central empirical argument in favour of the sociology of scientific knowledge. This normative character, however, is also incorporated in analyses of distributed cognitive systems: a distributed cognitive system has some functions, and each particular effect or output can be qualified as performing, or not, the system's function; i.e. the output of a cognitive system achieves, or not, the system's cognitive goals. It is difficult to specify what exactly the function of producing mathematics is. One reason of this difficulty is that there is, in fact, no such a specification. The general goal of the

CSM is not immune from historical and cultural variation. Saying that the CSM has changed its goal means that what was taken as the proper product of the CSM yesterday would now appear as being mistaken, i.e. as not fulfilling the CSM's function. In some cases, indeed, the evaluation does not merely question the truth and suitability of some particular results, but casts doubt on the very nature of the enterprise. This is what happened when mathematicians and philosophers questioned the legitimacy of computer-dependent proofs. It is often mathematicians who elaborate representations of what mathematics is and ought to be. With these representations, mathematicians evaluate the production of the CSM; or conversely, mathematicians evaluate the production of the CSM and elaborate representations, for post hoc rationalisation, of what mathematics is and ought to be. Thus, it is the very function of the CSM that is specified by mathematicians themselves. The CSM is made functional only to the extent that mathematicians provide it with a specific function. When mathematicians decide that the 4CT is actually a theorem, by the same token they also specify the function of the CSM as including the production of computer dependant proofs. Barnes' bootstrapping process is at work here: the representations of the nature of mathematics constrain the production of the CSM and thus (thought partially and indirectly) construct its own referent. Bloor (1976, chap. 6) shows variations of reference in the history of mathematics by exhibiting 'alternative mathematics', i.e. proofs and ways to reason that would not satisfy our contemporary standards. Bloor emphasizes the distinctive features of the mathematics of, among others, Greek arithmetic and their notion of number (as based on the practice of counting *versus* the practice of measuring which took some importance in the sixteenth-century technology), and the 'calculus' of the seventeenth-century (as, e.g. Cavalieri cancelling infinite quantities in $(b.1/2.\infty).h/\infty$ *versus* using limits). The case study of the 4CT illustrates well the point, since the question of whether the 4CT is a theorem and belongs to the corpus of mathematics or not is a question about which standards to apply when deciding what counts as mathematics. Tymoczko has well shown that at stake are the essential characteristics of mathematics and the very definition of the discipline. Analysing what Bloor qualifies as alternative mathematics with the framework of distributed cognitive systems makes explicit the fact that what is beyond the variation of "cognitive styles" (Bloor, 1976, p. 110) is not different minds with different capacities, but different means for cognition that serve different goals. In the end, what is important is that the production does contribute to maintain the social structure, and it does so through the distribution of an evaluative representation of mathematics. If people believe that doing mathematics is worthwhile, then they will devote their own resources, either directly or through funding — the discipline of mathematics is then maintained, a positive feedback process is in place.

11.4 Rethinking the new 4-colour problem

11.4.1 Methodological consequences for the history of mathematics

The new 4-colour problem is that of knowing what is the status of Appel and Haken's computer dependent proof. Tymoczko's (1979) philosophical analysis has pointed out the deep differences between computer dependant proofs and more traditional proofs. His work is highly relevant for the cognitive studies of mathematics because his analyses bear on the cognitive practices of mathematics and because he considers proofs as external objects with which mathematicians can think. The 1979's article can thus be read as an early contribution to the study of situated cognition and knowledge in action (Knorr Cetina, 1999). In particular, proofs are artefacts that can be surveyable or not. Accepting computer dependent proofs implies abandoning the surveyability of proofs; it implies abandoning a cognitive property of proofs and limiting the aspect of mathematical cognitive practice that consists in surveying proofs. MacKenzie (1999) provides a historical analysis of Appel and Haken's proof and its reception as a key event in the transformation of mathematical practices. How are such changes in mathematical practices brought about? MacKenzie insists on the social aspects of these transformations by pointing out that actors of these changes are put in front of genuine choices, with no pre-established rational recipe to decide. This is shown by exposing the alternatives and the arguments advanced by the participants to the debate as both reasonable and grounded. Yet, MacKenzie's historical account does not pin down the social motives that explain participant's choices. In particular, he does not show that those who engaged in computer dependent proofs had special interests in introducing computers in mathematical practices others than the advance of mathematical knowledge (epistemic interest, would say Giere), which is - one could assume - common to both sides of the debate. In fact, MacKenzie (1999) does not claim to explain why mathematical practices changed. His article is just pointing at the existence of two positions with regard to the 4CT, thus showing an opening for sociological investigation. In the absence of a causal account, however, a teleological account is still an easy temptation: One could say that the development of computer dependent proofs was really the inevitable rational development for mathematical practices, and that the opponents had social non epistemic motives, as the fear of being outdated and loosing one's competitiveness, the fear of a technology that is not mastered. Indeed, Appel and Haken's work has been massively acclaimed, while the critics are few and sound backward thinking when assimilating what mathematics should be with what mathematics had been. The impression of inevitability is increased because the discourses analysed by MacKenzie seems to have played little role in determining the actual practice: Philosophical debates around the nature of the 4CT seem to have been just that: talks about a conceptual problem (how to qualify Appel and Haken's work) rather than determinant of mathematical knowledge and practice. Even the mathematicians most opposed to computer dependent proof did not claim that Appel and Haken's result was false, and they did not ask mathematicians to stop working with computers. The argument was essentially about the status of Appel and Haken's work, and the necessity to continue to search for a more traditional proof. Discourses on the legitimacy of computer dependent proofs arose as *post-hoc* rationalisation with no real consequences on actual practices. This situation contrasts with other case-studies in the sociology of scientific knowledge, where the discourses of the proponents of alternative theories aim at determining and do determine the development of scientific knowledge. Such discourses are knowledge in the making, and studying the social and psychological conditions out of which they arose is studying the conditions of scientific knowledge development. Discourses about the 4CT, by contrast, belong more to the philosophy of mathematics than to the development of mathematics itself. So we can question the extent to which they reveal something about mathematical practices and the evolution of mathematics. The sociologists' work, however, is to dig further into the determination of the changes, rather than stop here. How can the historian accounts for the historical importance of the 4CT? What is the relevance of the debate between the enthusiasts and the sceptics of the 4CT? Is the new 4-colour problem only a development in the philosophy of mathematics, or does it really talk about a key phenomenon of the history of mathematics? The notion of evolving distributed cognitive system comes to the rescue for the naturalistic understanding of the historical significance of the 4CT. The first thing that the framework does is to enable the description of the stake of the historical event. The stake of the new 4-colour problem is the organisation of cognitive labour for mathematical knowledge production. The question does not primarily bear on the essence of mathematics, or on what sort of knowledge qualify as mathematical knowledge; the question more pragmatically and importantly bears on mathematical practices, it bears on the social organisation of an academic community, it bears on the distributed system of mathematical cognition. Appel, for instance, describes his work with the computer as complementing his own work of mathematicians: the computer "was much more successful [on its task of checking reducibility], because it was thinking not like a mathematician." He also had a clear idea of what he could expect from the computer — its domain of competence. Appel explains: "In the construction of our unavoidable set we constantly made minor modifications in the construction routine to avoid difficulties caused by likely to be reducible configurations we could not prove reducible by computers." It means that even though he believed some configurations were truly reducible, Haken and Appel did not expect the computer to be able to prove it, and adapted their strategy consequently. The introduction of computers in mathematical practice is therefore an evolution of a distributed cognitive system. Computers are trusted for the production of types of representations and are thus given a specific place in the cognitive architecture of the CSM. In the previous chapters, I argued that changes in distributed cognitive systems can be explained by the distribution of representations attributing new cognitive functions, which are representations of trustworthiness. The method for the scientist of science is therefore to track down representations of trustworthiness, account for their distribution and their causal role.

11.4.2 How mathematicians came to trust computers

Giving a role to somebody or something in a distributed cognitive system is to trust that this person or thing will perform its task satisfactorily. Trusting behaviour has a very bad press in mathematics. It is thought that mathematicians are not allowed to trust anything but their own reasoning. For instance, it is thought that working out a theorem is not just understanding it, but also understanding why it is true — i.e. checking the proof for oneself. Surveyability in mathematics is also a practice based on the idea that mathematicians should not just trust each other, but get their own persuasion after having surveyed the proof. Yet, trusting behaviour does occur in mathematical practice. Surveying a complex proof can itself require sharing the tasks, and thus trusting those people who checked the parts of the proofs dealing with the topic they are experts in. As soon as you have specialisation and distribution of cognitive labour, then trust is involved. Trusting is a mathematical practice among others. With the advent of computers, mathematicians could easily wonder: what should we trust them for? As word processors? As heuristic tools for graphic generation, or for suggesting proofs? As inescapable partners in proving? The first use gets an unquestioned yes; the second use of computer is already controversial and need may specification about when to use and how to interpret the resulting, computer generated, representations; the third use of computer is marked by a historical event, the proof of the 4CT. The 4CT epitomizes the problem of trusting cognitive artefacts. That the organisation of mathematical cognition is really a question of what to trust and for what, is well illustrated in the history of the acceptance of the 4CT. It is here that MacKenzie's social account shows that there has been different views to be accounted for:

> For some, like Bonsall, to put one's trust in the results of computer analysis is to violate the very essence of mathematics as an activity in which one's own human, personal understanding is central. To others, such as the 'core group' working on computerized reducibility proofs, using a computer is no different in principle from using pencil and paper, which is of course universally accepted in mathematics (MacKenzie, 1999, p. 50)

Trusting behaviour, and especially trusting behaviour in mathematics, is not the mark of unquestioned gullibility. The trustworthiness of the people and cognitive artefacts that are trusted has been assessed. For Hardwig, for instance, A must know that B is trustful (i.e. honest in this communicative situation), competent, conscientious, have adequate epistemic self-assessment. "The reliabity of A's belief depends on the reliability of B's character." (Hardwig, 1991, p. 700). This is transferable to the reliability of cognitive artefacts. Tymoczko (1979) presents the following thought experiment: Martian Mathematics takes the sentence "Simon's says p" as a proof of p. This can be justified because Simon was an incredibly good mathematician, who stopped giving proofs of his assertions in order to spare time. Martian came to trust Simon on the basis of his proofs, then they trust him just because he was Simon. Tymoczko concludes that "the appeal 'Simon says' is an anomaly in mathematics; it is simply an appeal to authority and not a demonstration" (p. 72), and he notices a few lines below:

> Computers are, in the context of mathematical proofs, another kind of authority. If we choose to regard one appeal as bizarre

> and the other as legitimate, it can only be because we have some strong evidence for the reliability of the latter and none for the former. Computers are not simply authority, but warranted authority.

This is a move towards classical 'social epistemology': beliefs acquired from testimony are justified if one has some knowledge of the trustworthiness of the source of information. What is still original today in Tymoczko's observation is the application of this piece of social epistemology to mathematics and cognitive *artefacts*. In the case of the 4CT, the reliability of the computer is represented by the belief that "[t]here are very good grounds for believing that this computer work has certain characteristics, e.g., that it instantiated the pattern of a formal proof of the reducibility lemma." (p. 73). Thus, assessing whether a computer is trustworthy requires the possibility of interpreting its output as showing that there exists a formal proof of so and so. This is persuasive only if the existence of a formal proof provides a sufficient ground for believing as true what the formal proof proves, so formalism in mathematics has its role in the assessment of computers' trustworthiness. The assessment of computers, in any case, is by itself a cognitive practice. MacKenzie (1993) explains the different means that have been used to assess the trustworthiness of computers, especially when their failure in performing the task could issue in great loss. Traditionally, he says, the confidence we have in computer systems has been empirically based, but new approaches, employing deductive means, have been used to check the "correctness of designs and programs." MacKenzie shows the important stakes and richness of means employed to assess computers' trustworthiness. For instance, "In the early 1980s, the US Department of Defense set out its Trusted Computer System Evaluation Criteria, known from the colour of the cover of the document containing them as the 'Orange Book"' (p. 52). In the case of the 4CT, however, the reliability of the computer was assessed empirically. Is this method sufficient? John Slaney, a specialist in automated theorem-proving remarks: "on the score of reliability computers have it over graduate students by a wide margin." (quoted by MacKenzie, 1999, p. 50). However, the trustworthiness of computers came to be seriously doubted when, in 1971, it was found that some original run of Durre's programme erroneously asserted that the "horseshoe" configuration was D-reducible. Appel and Haken's computer's results was checked by comparing results between runs of computer programmes. In the refereeing of their paper, Appel and Haken have had their computer's results compared with Allaire's results with his own programme: he found 400 agreements versus 0 disagreements and concluded that the programme could not include a bug. But a year later the comparison was done with 2669 reducible configurations and 2 disagreements were found; a re-run of the programme, however, issued no disagreement.

The argument is that the 4CT is a key event in the evolution of the CSM, which is itself analysed in terms of the distribution of representations of computers as trustworthy for certain cognitive tasks for the production of mathematical knowledge. The 4CT largely contributed to the distribution of such representations, and the distribution of these representations was rendered possible because (1) people interested in the introduction of computers in mathematics where so because (among other things) they convinced themselves, and could convince others, that computers were trustworthy for the tasks imparted to them — the evidence was framed in terms of 'empirical tests' of reliability, and later on, with formal deductive means, and (2) beliefs about the trustworthiness of computers was brought up through inferences involving the premises 'a theorem is true if there exists a formal proof of that theorem,' and 'computers can implement formal proofs.' The first premise sends back to the history of formalism in Mathematics, the second premise sends back to the history of computer science and its long lasting and promiscuous relations with mathematics. To which extent, indeed, were computers *designed* to do mathematics? Computers would not have had a role in mathematical cognition if they were not thought as 'fitting in'. Such thoughts are not discoveries popping out of nowhere: computers have always been thought with and through mathematics. In the twentieth century, both mathematicians and computers were thought as manipulating symbols through syntactic and logical rules. The analogy went even further with the advent of AI, where human cognition was primarily thought on the model of the mathematicians manipulating symbols. There was therefore‘naturalness’ in implicating computers in mathematical cognition. This naturalness stemmed from an already existing distribution of representations of computers as mathematicians of a sort. One way to explicate conceptual revolutions is to show that there were not so revolutionary after all, which simply means that the revolution did not arise from nothing. In the case of the 4CT, this explication is appropriate: there is continuity in the distribution of representations of computers as trustworthy for doing mathematics: these representations were present before the 4CT, and before Heesh’s first appeal to computer means. They were distributed in the population of mathematicians for reasons that are independent of the 4CT. Then, the 4CT *contributed* to the distribution of the representation of computers as trustworthy for doing mathematics. This is due, in particular, to the fact that “every use of an instrument or apparatus testifies to its reliability and hence to the standing of its inventors or current guardians” (Barnes et al., 1996, p. 115). The post-4CT saw an increase in computer use within mathematics, this increase was continuous and motivated by several factors, including factors independent from the 4CT, such as increase of computer power and usability. Moreover, computer dependent proofs are still rather few, they include the proof that there are no finite projective plan of order ten (published in 1889) which involved examination of 10^{14}cases, and the solution of the Robbins Problem (published in 1997), which can, however, be checked by human mathematicians. Finally, the most important continuity pre- and post-4CT concerns the practice of distributing mathematical cognition. This practice, as argued in the previous section, is since the beginning part of what is done when one does mathematics. Edward Swart draws upon the fact that the practice of mathematics essentially involves using artefacts for advocating computer assisted theorem proving. He says:

> [T]here is precious little substance in the world of mathematics that can be done... without some assistance from pencil and paper. And for the most part I regard computer-assisted proof as just an extension of pencil and paper. I don't think there is some great divide which says that OK, you are allowed to use pencil and paper but you are not allowed to use computer because that changes the character of the proof... I find such an argument strange (Edward Swart, interviewed by A. J. Dale, Ontario, 1 June 1994. Quoted by MacKenzie, 1999, p. 50.)

Swart is right in drawing our attention on the continuity in Mathematical practice. But of course, his argument hides an important change: there is indeed a divide between the use of pen and paper and the use of computers, and the cognitive functions attributed to these two artefacts are different in important respects: the problem raised with computer assisted proof is not the use of computers as word processors. To which extent was the organisation of mathematical knowledge production changed in order to integrate computers? Computers would not have had a role in mathematical cognition if they did not already somewhat 'fit in' — as artefacts for doing mathematics. But they would not have had a role either if people - such as Appel and Haken - did not actively ascribe a cognitive function to them. Analyses of distributed cognitive systems show that technological innovations lead to technological change only when the new technologies are given a cognitive function into distributed cognitive systems. The ascription of cognitive functions is not necessarily the result of actions by people consciously designing distributed cognitive systems (see chap. 9). It is much more often the result of tinkering with actual opportunities. Thus Haken explains his choices of using computers for proving with the following sentence: "If you run into terrific complexities, do not go on, but look for more powerful means ... stronger tools of higher mathematics" (Wolgang Haken, interviewed by A.J. Dale, quoted by D. MacKenzie, 1999). Stronger tools have often meant in mathematics the development and use of new mathematical theories, but it can as well refer to artefacts.

Computers were the tools in sight for doing mathematics. The importance of the familiarity one has with the potential cognitive artefact is not only a factor of innovation, but also a factor of acceptance of this innovation. Thus, according to Tymoczko (1979, p. 81):

> ... placing [the 4CT] in a historical perspective can be very illuminating. I suggest that if a "similar" proof had been developed twenty-five years earlier, it would not have achieved the widespread acceptance that the 4CT has now. The hypothetical early result would probably have been ignored, possibly even attacked [...]. A necessary condition for the acceptance of a computer-assisted proof is wide familiarity on the part of mathematicians with sophisticated computers. Now that every mathematician has a pocket calculator and every mathematics department has a computer specialist, that familiarity obtains. The mathematical world was ready to recognize the Appel-Haken methodology as legitimate mathematics.

Putting all these efforts in computer assisted proofs could not have been done with the serious confidence that computers could perform the required task. The confidence was achieved by a sufficient familiarity with computers, and several representations about their trustworthiness in what concern the production of formal proofs.

For Tymoczko (1979), the 4CT is a change in paradigm. Tymoczko rightfully insists on the important discontinuities between traditional proofs and computer assisted proofs. His work is essentially the one of a philosopher of mathematics spelling out the consequences of Mathematicians' actions and choices. From a historical and sociological point of view, however, there is important continuity in the distribution of practices and ideas among mathematicians. The representations attributing cognitive functions to computers within mathematics sustained the change in mathematical practices, and their spread is accountable with different factors, including prior beliefs about mathematics and prior beliefs about computers. This also shows that the history of the philosophy of mathematics is part of the history of mathematics, since beliefs about mathematics determine what mathematics is. Thus, giving to the 4CT the status of a theorem is opening mathematical practices towards computer assisted proving. Labelling Haken and Appel's 1976 paper a proof is not just an act of classification, it is also setting the boundaries of the distributed cognitive system of Mathematics. If mathematicians say that computer dependant proofs are part of mathematics, then we must include computers in the CSM. This holds for research managers and head of departments of mathematics, who now provide adapted computer facilities to mathematicians, and this hold for the analyst of distributed cognitive systems.

Computer use in mathematics has yet another consequence on the borders of the CSM: the need for expertise in computers redefines the boundaries between computer science and mathematics. In the 60′ mathematicians were knowledge providers for computer scientists. With the 4CT, mathematicians become users of computer science's knowledge. At the same time, the field of computer science gained more and more autonomy with regard to mathematics, in such a way that we can say that there are interactions between two distinct distributed cognitive systems. The boundaries being crossed often and at several points become fuzzier. There is interdisciplinarity. The dependence of mathematicians upon computer scientists raises further questions, as changes in distributed cognitive systems bring changes in social organisation. Along these lines, a new rendering of the new 4-colour problem consists in seeing the advent of computers in mathematical distributed cognition as a process of mechanisation of cognitive labour. The labour involved in proving involves a mode of production, a social organisation, which differs from traditional (pre-mechanised) mathematics to computerised mathematics. The analysis of the distribution of cognitive labour directs attention to the advent of computer scientists as new experts for mathematical knowledge production: computer scientists. The first programmer involved was Karl Duerre, a secondary teacher at the time of his enrolment, who did his Ph.D. in Mathematics on methods used in the proof of the 4CT. Duerre, however was sufficiently expert in computer science to programme in Algol 60 and punch the data on punch cards. The second programmer for the 4CT was

Koch. Koch was a graduate student in *computer science*[6] when he started to work with Appel and Haken on the 4CT. He thus had the chance to co-sign the first part of the 1976 paper of the 4CT. Most importantly, Appel had a good expertise in computing. Appel's speciality was mathematical logic, the topic most directly linked to computing. He also had taken computer courses and had written several large programmes. Both Koch and Appel were sufficiently proficient so as to programme in assembly language, which gives more efficiency. *Cognitive dependence*, says Hutchins, lead to social dependence. With the mechanisation of proving, mathematicians become dependent on programming experts. While the new social organisation settles, so does a discourse of justification legitimating the new social order: it is the discourse about the nature of mathematics. Also, mathematicians become dependent on the owners of computers as the new means of knowledge production: a lab in nuclear physics at Brookhaven and the administration of the University of Illinois in the case of the proof of the 4CT. Do we have here a capitalisation in the domain of knowledge production? The question is not just rhetorical: in a domain of biology such as the sequencing of the human genome, or in biotechnology, capitalisation raises genuine worries. In the computer domain, capitalisation raises worries with regard to the management of information on the internet, especially concerning the powerful companies owning search engines. In education, including education in mathematics, computer equipment have arised to the status of essential pedagogical means, up to the point where computer facilities provide a good criterion on the wealth — and consequent quality — of educational institutions.

Conclusion : the new four-colour problem, I argued, is not primarily a problem about the nature of mathematics. Historically, it is a problem about the design of a distributed cognitive system, of which philosophical questions about proofs are just an aspect. How did mathematicians come

[6] Departments in Computer Science flourished at the beginning of the 6o'. The University of Illinois at Urbana opened its first graduate degree programme in 1966.

to introduce computers in their cognitive practices? The answer required investigating the distribution of representations of the trustworthiness of computers for mathematical tasks and the causes of this distribution. The evolution of the CSM has not been the result of an agreed upon convention about the nature of mathematical proofs. The evolution resulted from the tinkering of mathematicians with tools at hand, computers in our case, that were found appropriate for achieving their own goals. Proving the four colour conjecture happened to be among the most prominent goal, and computers where found to perform the task imparted to them.

It might be objected that the analysis in terms of distributed cognition could not grasp essential features of Mathematics, such as the fact that some theorems are important or beautiful and others not. Indeed, theorems, be they Whiles' proof of Fermat's conjecture or some trivial proposition, are equally outputs of the system and cannot be differentiated as such. My answer to the objection is that an analysis in terms of distributed cognition may not explain everything about Mathematics. It is a non-reductive theoretical framework. So the aesthetics linked to Mathematics may be accounted for in terms of, say, cognition and emotions. Yet, the analysis in terms of distributed cognitive systems can provide better historical accounts and new insights into the nature of mathematical knowledge. The problem of accounting for the difference in importance of mathematical results could be analysed in terms of how much processing the required to output the result, or it could be said that the importance of a mathematical result depend on its incorporation as a theoretical tool of the CSM for the further production of mathematical knowledge. The notion of distributed cognition is most appropriate when analysing the introduction of artefacts for achieving cognitive tasks. The history of mathematics include important evolutions of the roles attributed to cognitive artefacts: for instance, graphs were first considered as trustful means for arriving at true mathematical knowledge, then their cognitive function was more and more restricted, from elements of proofs to heuristic means to manipulate with care. The compass and the ruler have had their fate tied to that of graphs: they were given a great role in processes of proving in geometry.

But this is not the case anymore. The causes of evolutions include inventions of artefacts, trust ascription, changes in methods and the evolution of beliefs about mathematics. Most importantly, I argued that the evolution of distributed cognitive system depend directly on the distribution of representations ascribing cognitive functions to elements.

Part IV

The exploitation of cognitive abilities and tools

Chapter 12

An integrated causal model for science studies

The objective of this thesis was mainly to investigate possible theoretical foundations of an integrated causal model for science studies. I have picked up theories from social studies of science, from cognitive anthropology and from cognitive psychology and shown that these theories can together form a solid and coherent ground for integrated and causal studies of science. The project of forming an integrated causal theory of the evolution of science can be tracked back to the 60's, when Campbell elaborated his "evolutionary epistemology." In the following section, I consider Campbell's project and critically assess the theory put forward with more recent theories of human cognition and cultural evolution. I argue that my own choices of theories — the epidemiology of representation, a Strong Programme approach, and a massive modularist view of the mind — form an optimal updating of Campbell's evolutionary epistemology. In the first part, I present and criticise evolutionary epistemology. I then present my own proposal as a way to update evolutionary epistemology with current relevant theories. I advocate: (1) dropping the methodological constraint of looking for processes of blind variation and selective retention at the expense of other constructive processes and mechanisms of knowledge production, but (2) retaining the integrative point of evolutionary epistemology, which implies taking seriously the results of evolutionary psychology. This sets a research programme in cognitive psychology of science, viz. understanding scientific cognition as implemented by massively modular minds. However, I argue that understanding scientific cognition requires studying the social embodiment of cognition as well as its biological implementation in scientists' brains. This gives me the occasion to sum up some of the points in the thesis, with a view on what they may bring to cognitive science in general.

12.1 EVOLUTIONARY EPISTEMOLOGY

Campbell's "Evolutionary Epistemology" is a research programme that fits the Integrated Causal Model: Campbell (1974)Campbell introduces it as a "descriptive epistemology" that "would be at a minimum an epistemology taking cognizance of and compatible with man's status as a product of biological and social evolution" (p. 413). Evolutionary Epistemology aims at providing a causal history of scientific knowledge that not only accounts for the human history of science making, but also includes accounts of the cognitive processes at the basis of this history and of the evolutionary history of the cognitive abilities implementing these cognitive processes. Evolutionary epistemology is therefore an integrated research, which spans biology, evolutionary psychology, cognitive psychology, sociology and history. For instance, Campbell, following Konrad Lorenz, advocates the understanding of Kant's categories of perception and thought as evolutionary products (1974, sect. 5). Yet, within this integrative perspective, there are a number of points of disagreements between Campbell's proposal and the one I would like to defend. Campbell, indeed, further asserts that there is one single principle at work at the levels of natural history, thought processes and science history: it is the principle of "blind-variation-and-selective-retention" generalised from Darwin's theory of natural history so as to account for creative thinking and the history of science. Concerning the history of science, Campbell fully takes on Popper's account of the "Logic of scientific discovery" and its principle of "conjecture and refutation". Concerning creative thought, Campbell (1960) develops his own argument, which puts at the centre stage of creative thought the "eureka" phenomenon. Concerning cognitive abilities, Campbell applies evolutionary biology to the cognitive apparatus, elaborating thoughts much akin to contemporary evolutionary psychology. In this chapter, I defend alternative accounts of thought processes and socio-historical development of science. While Campbell based his integrated model of scientific development on the single principle of blind-variation-and-selective-retention, which would account for natural history, the dynamic of thought and the history of science, I argue that different processes are at work at each level and that Darwinian Theory does not necessarily apply to scientific cognition and scientific development. The idea is that while integration requires showing how the biological, cognitive and historical explanations match and combine into a single more exhaustive one, there is no need to assume that the explanatory blocks, accounting respectively for natural history, cognition and social history, are of the same type. Indeed, I will point out that current theories in sociology and cognitive psychology describe mechanisms for the production of knowledge that differ from blind variation and selective retention. The conclusion is that the Darwinist model of evolution applies to the evolution of epistemic mechanisms (EEM) of the structure of the brain, but do not extend to an Evolutionary Epistemology of Theory (EET) (typology introduced by Bradie 1986. EET takes that the evolution of scientific theories is based on blind variation and selective retention; it hypothesise that the analogy between biological evolution and the evolution of scientific theories can reveal deep similarities. There are two problems with EET: the first is blind variation, and the second is selective retention. Rather than blind variation, I argue that cognitive processes are processes guided by the domain specific cognitive abilities that characterise the human mind (i.e. human epistemic mechanisms), the cognitive principle of relevance, and the affordances constituted by the distribution of representations in the scientists' minds (acquired knowledge) and in the environment. Rather than selective retention, I argue that there are several, diverse, social and cognitive mechanisms that determine how representations stabilise in the scientific community—the role of cultural attractors and of social institutions have been analysed above (parts II and III). In brief, the alternative account I advocate is an epidemiology of scientific representations. The epidemiology provides a framework for the naturalistic study of scientific evolution, but it does not specify *a priori* the mechanisms through which evolution occurs. It enables taking into account psychological theories as specifying the mental mechanisms through which scientific representations are produced and processed, and sociological theories as specifying the social mechanisms that affect the distribution of scientific representations. For Campbell, blind selection and selective retention is a necessary process of evolution: evolution implies the generation of genuinely new items, which means that the generative process cannot be biased by the value of the items (in terms of fitness); the generative process does not embed knowledge of the value of the new items. As an analytical truth about evolution, or as an abstract principle that can always describe, at some level, the processes of evolution, there is nothing to say against blind variation and selective retention. But when one attempts to explain the detailed causal processes through which evolution takes place, then, blind variation and selective retention is an insufficient analytical tool. This point is already well known by evolutionary biologists, who have described many processes through which evolution may have occurred, including drifts, spandrels and exaptation. The criticism of evolutionary epistemology I develop in this section is a means for pointing out the richness and diversity of the social and cognitive processes out of which scientific knowledge is constructed. In the spirit of evolutionary epistemology, one goal is to integrate the results from evolutionary psychology, psychology of science (including psychology on creativity), and sociology of science. But this integration is hindered by the further attempts to impose the Darwinian model on all processes, at all levels, of knowledge making. This modelling constraint tends to hinder rather than foster research.

12.1.1 Blind variation

Blind selection and selective retention require a decoupling of variation and selection. The generation of beneficial items is not more probable than the generation of non-beneficial items. It is in that sense that variation is blind: it does not 'see' in advance whether the item generated will be selected or not. Campbell asserts that the generation of new ideas is accountable in terms of blind variation and selective retention. The Darwinian process is intended to account for the creativity of scientific thinking. But are psychological processes of scientific belief formation based on blind hypothesis formation [1]

One important problem with Campbell's thesis on cognitive processes relates to the cost in time and energy of the blind search he hypothesises as being at the basis of thought. As Campbell himself notes, blind search implies an enormous number of possible thought-trials to be searched before one can select a solution. The tremendous number of non-productive thought trials that a blind-variation-and-selective-retention necessarily produce makes the cognitive system unfit for survival, where decisions need to be taken quickly (e.g. when facing a predator) and where energy resource is rare and scarcely allocated (although the brain is a high consumer). Moreover, this does not correspond to the recent findings of cognitive psychology: decisions are actually taken quickly, relying on innate knowledge (such as naïve physics or naïve psychology and on fast and frugal heuristics) that guides reasoning. Campbell (1960), however, had considered these counter-arguments. One strategy he adopts is to point out that blind-search-and-selective-retention is not that much time and energy consuming because it functions with a simple stopping rule for the search: being selected when answering some criteria. This stopping rule contrasts with the ones of unbounded rationality, which requires that one gets at the best solution, and optimisation under constraints, which

[1] see Kronfeldner (to appear) about arguments on the compatibility of Campbell's view with the view that hypothesis formation is biased; my argument follows some of her points, but I focus on blind-variation-and-selective-retention as a tool for describing the cognitive processes themselves

requires that one gets at the best solution available given one's cognitive constraints. Thus, Campbell escapes much of the criticism that the advocate of bounded rationality address to traditional theories of rational thinking. Campbell refers to Newell et al. (1958); he knows what is at stake (e.g. problems with informational explosion) and acknowledges the credibility of the heuristic approach. Campbell's stopping rule is an instance of Herbert Simon's satisficing process. Furthermore, Campbell (1974) Campbell allows its system to incorporate "shortcuts" to full blind-variation-and-selective-retention process, thus making a nested *hierarchy* of selective-retention processes (1974). Domain specific heuristics, innate knowledge or Kantian categories are such shortcuts because they allow compiling the solution without blind-search or limit the blind-search to a restricted domain. Campbell, however, quickly points out that (1) such cognitive abilities are themselves produced through blind-variation-and-selective-retention and (2) "such shortcut processes contain in their own operation a blind-variation-and-selective-retention process". Within the perspective of evolutionary psychology, the first point is granted, to the extent that the cognitive processes result from evolved cognitive abilities (but I still question whether learning relies on blind variation and selective retention). The second point, on the other hand, is at odd with much of recent theories in cognitive science. For instance, the heuristics described by Gigerenzer & al.(1999), with the exception of the satisficing heuristics, operate when all the possible choices are available to the decision maker, therefore not including blind search. The cognitive processes function with a pre-established set of cues that are taken as sufficient for making 'rational' decisions. The naïve theories hypothesised by development psychologists likewise do not include blind search: they operate when triggering conditions are met, and their operations are fully specified. Sperber and Wilson's (1986) account of the cognitive processes for verbal understanding do function with a satisficing procedure, but the search that precede stopping is not blind: it is guided by the structure of the common cognitive environment of the speaker and the audience and the communicative principle of relevance. These examples, of course, do not show that cognitive processes never operate on blind variation. They show only that blind variation is not at the centre-stage of cognition; it is not, as Campbell would have it, the core principle of thinking. As a consequence, there is not a single principle at the basis of both biological evolution and cognition. The two levels must be distinguished so that the particular functioning of each can be analysed.

Campbell is misled by the examples he takes as paradigmatic thought processes because he heavily relies on scientists' intellectual discoveries and their phenomenological account, such as the Eureka phenomenon and Poincaré's essay on mathematical creativity. But according to Campbell's own emphasis on the cognitive apparatus as an evolved organ, scientific inventions can hardly be taken as paradigmatic of cognition in general: the cognitive apparatus evolved to cope with day to day needs and dangers. The human brain, in particular, has evolved when the human specie was hunting and gathering and our cognitive apparatus is therefore designed for coping with the tasks of the hunter gatherer as performed in the manner of our ancestors. Science, on the other hand, is a very recent cultural achievement; science making cannot be a biological function of the human brain because the ability to do science is too recent for being included in our biological evolutionary history. Taking evolutionary psychology seriously requires that the theories of cognition—including scientific cognition—be compatible with some evolutionary history of the biological function of the cognitive processes. Whence Gigerenzer & al.'s (1999) emphasis on fastness and frugality, which provide obvious advantages in the face of natural selection; whence also the emphasis on the domain specificity of cognitive processes. The conclusion of these researches is that the mind is constituted of many heuristics that solve problems in specified domains; it is an "adaptive toolbox". In comparison, it is therefore implausible that blind-variation-and-selective-retention evolved as a domain general cognitive process, on top of which "shortcuts", such as heuristics, would further evolve. Evolutionary psychology re-centre the investigation of cognition on real-world tasks rather than on abstract problem solving (such as scientific theorisation) because it requires assessing the adaptive behaviour enabled by the cognitive processes. The assertion that the biological functions of cognitive processes are designed (through evolution) for coping with the environment (so as to ensure survival and reproduction) leads to the investigation of "ecological rationality" as a property of cognitive processes Gigerenzer et al. (1999). Evolutionary epistemology, by its very definition, must be compatible with the above findings of evolutionary psychology. Rather than the scientists' discoveries, it is the ability to solve problems present in the environment that determined the selection of the genetic basis of human psychology that is best likely to give us the key of evolved cognitive abilities. It then appears that it is little probable that the evolutionary history of the human cognitive organ would have constructed a cognitive device that implements blind variation of mental representations and selective retention of these representations. This goes against Simonton's account of creativity (1999), which state that hypothesis formation is based on a subconscious random generation of ideas: only selected ideas come to consciousness, but a massive number of unconscious random ideas have been previously generated. In addition to its low adaptativeness (the generation of a massive number of random ideas seems too costly for being selected by natural evolution), there is little empirical evidence in favour of hidden, unconscious, chaotic generation of ideas (see Sternberg 1999).

How can we pass from ecological rationality to scientific rationality? The latter is oriented towards the discovery of truth rather than towards gain in fitness (see section 5.4). The passage is done through communication and the social aspects of knowledge making. The fact that communication and social interaction constitute an essential part of scientific practice is nearly a truism. Scientific cognition is oriented towards social interaction, and in particular, towards the communication of new ideas, whose appeal is importantly dependant on their being *taken as true* by the audience. Scientific cognition aims at communication, and so it aims at truth. The relations between scientific thinking and scientific culture (knowledge and practices) turn to the principles of scientific communication: scientists think about communicating new ideas, scientists' thoughts are determined by their understanding of the communicated ideas, scientists judge and critically assess what other scientists communicate, etc. The importance of communication in the social evolution of science is actually much present in Popper's epistemology. Commenting on Campbell's evolutionary epistemology, Popper (1974) emits a criticism, which he claims to be "related to the difference between man and animal, and especially between human rationality and human science, and animal knowledge". Popper's point stresses the argumentative practice that is at the heart of science and that makes criticism possible. In doing so, Popper points out that science is a social practice that involves people communicating and judging their communications. It is this fact that put the problem of truth and scientific rationality back into scientific cognition. With regard to truth, Popper says: "I think that the first storyteller may have been the man who contributed to the rise of the idea of factual *truth* and *falsity*, and that out of this the ideal of truth developed; as did the argumentative use of language". The ideal of truth and the practice of argumentation are therefore stemming from social interactions; they are constitutive of scientific cognition because science is a social activity. On this basis, new constraints on scientific cognition arise: scientific cognition must conform to the rules of scientific rationality, which is made of historically developed ideas about truth preserving cognitive processes. Through this complex path, going through social interaction, scientific cognition becomes rational in the normative sense, rather than ecologically rational (c.f. section 4.1.2).

In evolutionary accounts of science, both individual cognition and social processes are given due roles, but the complex relation between the two seems to be missing: it involves the cognitive processes underlying communication. The relevance model of verbal understanding does not include blind variation as a cognitive process, and it also differ from retention of content from one mind to another. Sperber and Wilson 1986 argue at length against the code model of communication, according to which content is coded into utterances by the speaker, and then decoded by the competent audience. Understanding, in particular, implicates constructing a mental representation of the speaker's communicative intention on the basis of linguistic input and some knowledge of the situation and what is known of it by the speaker. Giving too little importance to cognition specific for communication and social interaction leads Campbell to a dilemma. Either he adopts the views of evolutionary psychology and assumes that human cognition in general, and scientific cognition in particular, is ecologically rational—he then misses essential features of scientific cognition, which aims at truth and objectivity, or he adopts a scientific centred view of human cognition—he then abandons the vow to be compatible with theories of man as the product of biological evolution. Putting communication and its cognitive principles at the centre stage of the evolution of science is something that the epidemiology of representations can do, and that is missing from competing evolutionary epistemologies.

Serendipity does actually lead to scientific discoveries (see Roberts 1989 for a set of historical examples), but is it really a key aspect of scientific cognition? Is hypothesis formation based on blind variation? Blind variation, in this context, does not mean that any variation can occur with equal probability, but that the chance that a particular scientific idea will occur is not influenced by the factors that determine its selection. So blind variation can be strongly biased, but the bias is not sufficiently strong and constraining to dismiss the role of blind variation. The initial motivation for including blind variation into scientific cognition comes from Popper's arguments against inductivism: it is never sufficient to gather data for creating knowledge, the scientists have to develop new hypothesis for accounting for the data. Induction does not solve the problem of scientific creativity, "trial and error" does. Sociologists of science have, to say the least, taken notice of the failure of inductivism, but rather than appealing to undirected variation or hazard, they have looked for further determination of the generation of ideas *outside* of the set constituted by data. Furthermore, these further determinations for the generation of scientific ideas determine both generation and the reception of these ideas. For instance, if a hypothesis is generated on the basis of an analogy with social structures, then it is likely that this same analogy will favour understanding and the perceived relevance of the hypothesis by the scientific audience; it will thus favour its reception. The set of possible constraints that affect both creation and reception goes well beyond 'pre-adaptations' or 'develomental constraints', which Stein and Lipton (1989) show to bias both biological and scientific evolution. The biases in variation are effects of previous cycles of blind variation and selective retention: there are vicarious processes, in the sense that they are processes produced by some lower level or more basic selection process (requirement about historical origin), and because they substitute the selective role the lower level process had before its origination (hypothesis about substitution). Admittedly, existing knowledge in science and existing biological state of affair do constrain, respectively, the generation of new ideas and the generation of new genetic combination. The point is that the remaining variation that makes up new knowledge is still not blind: it is guided by both ideas acquired from the cultural background and by evolved mental mechanisms. In the end, there appear to be a coupling of variation and selection such that blind variation cannot be said to properly characterise scientific creativity. At a minimum, the Darwinist framework seems, at this point, to hinder rather than foster research, as it unwarrantedly deny connexions between creativity processes and factors of reception. One point that strongly goes in favour of a connexion is that the reception of a new scientific idea depends on the understanding the communicated idea. But this understanding is itself a creative process, whose success is rendered possible because the audience have similar cognitive abilities and share the same background knowledge as the one expressing the new ideas. This constitutes a strong connexion between generation by individual scientists and selection by the scientific community.

12.1.2 Selective retention

According to the traditional view of evolutionary epistemology, blind variation as generating new ideas occurs within scientists' minds, while selective retention is mostly a social process involving scientists checking the work of others and choosing the best of it. Selective retention involves a process of selection that well describes the fact that not all of scientists' ideas gain the status of scientific knowledge and get distributed in the scientific community. But selective retention involves also a process of retention. Darwinian theory holds that retention is done through replication. In biology, it is DNA sequences that are replicated; in science, the replication is replication of beliefs or ideas... and the replication happens through communication.

David Hull, whose work can be understood as a refinement and updating of evolutionary epistemology (1988; 2001), specify what replicators are in the evolution of science:

> the replicators in science are elements of the substantive content of science — beliefs about the goals of science, the proper ways to go about realizing these goals, problems and their possible solutions, modes of representation, accumulated data reports, and so on [...] These are the entities that get passed on in replication sequences in science. Included among the chief vehicles of transmission in conceptual replication are books, journals, computers, and of course human brains. As in biological evolution, each replication counts as a generation with respect to selection [...] Conceptual replicators interact with that portion of the natural world to which they ostensibly refer [...] only indirectly by means of scientists. (p. 116)
>
> Conceptual replication is a matter of information being transmitted largely intact from physical vehicle to physical vehicle.
>
> In sum, conceptual replication is a matter of ideas giving rise to ideas via physical vehicles. (p. 117)
>
> One difference between biological and conceptual evolution is that in biology genes make genes ... conceptual replicators do not, on their own, produce copies of themselves. They do so only via their most important agents — individual scientists.

> Thus, in scientific change, scientists are the chief agents in both replication and interaction. However, on my analysis this difference is not sufficient to preclude a single analysis applying equally to both. (p.123)

Theory drift often happens in science, but this fact does not falsify darwinist account of the evolution of science, since replicators are smaller entities than theories. Theories are conceptual systems, and selection and variation operate on concepts. The problem is that replication at the conceptual level does not either properly describe the mechanisms through which representations are distributed and stabilised within a community. In order to make this point, I only briefly review the arguments advanced by Sperber (1996a), and Sperber and Claidière (2006) against Darwinian models of cultural evolution. The bulk of the argument is that replication is an unwarranted simplification of the complex socio-cognitive processes through which cultural phenomena arise. This is because:

> representations don't in general replicate in the process of transmission, they transform; and [. . .] they transform as a result of a constructive cognitive process. Replication, when it truly occurs, is best seen as a limiting case of zero transformation.

The consequence is that concepts or ideas are not replicating well enough to undergo effective selection: the rate of change is such that selection cannot be consequential on evolution. In place of replication and selection, Sperber appeals to the role of several factors of stabilisation of representations. Among those factors, importantly lies the rich and universal human cognitive endowment. For instance, a natural language is known and distributed within a population not only because children learn to speak on the basis of what they hear, but also because they have an unlearned ability to learn languages. Likewise, in chapter chap. 7, I have argued that the number sense has been a factor of both selection and development of representations in the infinitesimal calculus. The psychological factors are not to be considered as environmental factors of selection of representations:

they are factors because they are involved in the construction, rather than the selection, of representations. It is in this constructive processes that lay essential causes of change and stability. Cultural propagation is "achieved through many different and independent mechanisms, none of which is central and none of which is a robust replication mechanism" (Sperber and Claidière, 2006, p. 20). In particular, imitation is not the main mechanism of transmission, but only if "the notion is stretched to cover a wide variety of quite different processes." Thus, the observed macro-stability, as manifested by "relatively stable representations, practices and artefacts distributed across generations throughout a social group", does not warrant the existence of mental processes insuring the micro-heritability of cultural items. Again, theories in psychology and sociology about memory, imitation and communication show that high fidelity reproduction is the exception rather than the rule; "the micro-processes of cultural propagation are in good part constructive rather than preservative". The causes of preservation and propagation often lay in the fact that "constructive biases" are shared in a population: I mentioned the universal human cognitive endowment, but the common environment, as leading, in particular, to similar aspects in individuals' histories, also causes shared constructive biases. The shared constructive biases cause the emergence of cultural attractors: in spite of the fact that transmitted representations are different from one another, the representations do not drift away through added transformations to strongly dissimilar representations, but the constructed representations tend to gather around an "attractor". Consequently, Darwinian models of cultural evolution are unsatisfactory because "cultural contents are not replicated by one set of inheritance mechanisms and selected by another, disjoint set of environmental factors."

12.1.3 The layered construction of knowledge

The Darwinian model for thinking the evolution of science is certainly a rich source of inspiration and discovery. Hull (2001), for instance, draws on the Darwinian model for explaining social processes of competition and collaboration in the sciences. In the same way as inclusive fitness in biological evolution accounts for kinship altruism, in the sciences, scientists promote both their own work and the work of those that use their work. The works of scientists thus have "conceptual inclusive fitness." However, the Darwinian model has also strong limits, and giving too much weight to one single mechanism of evolution hinders rather than fosters cognitive and social investigation of the processes of cultural evolution. The sociology and history of science of these last decades have pointed out the social processes at work in scientific knowledge production. These include the institutional constitution of science, the coercive strength of scientific traditions (including the norms of rationality), the self-referring aspects of scientific beliefs, the goal-orientation of research, the role of trust in science, novice-expert interactions and how scientific practices are taught and learned, the reliance on external values and beliefs, the negotiations during scientific controversies. The abstract and methodological Popperian picture of conjecture and refutation is given more sociological reality, which implicate a complexification that cannot any more be grasped with the Darwinist process. Blind-variation-and-selective-retention seems, at this stage of sociological and psychological knowledge, not able to account for the social factors determining scientific practices, including scientific judgements, the forms of justifications, rebuttal and assent, and scientific creativity. Campbell's ambition to find a unique principle accounting for biological evolution, thinking, and scientific evolution provides an over-simplified picture of cognition and culture. The naturalisation of science studies passes first through an integration of cognitive and social studies of science. Imposing the Darwinist model on the evolution of science leads to bypass too much of the results in cognitive psychology and sociology of science.

cCampbell's BVSR in Heyes and Hull

objectivity in science.

nativism as a fact to take into account for explaining the evolution of culture.

One can distinguish several projects under the label of evolutionary epistemology: The most radical project is the application of the Darwinian model in order to account for the evolution of knowledge. I have argued that this project, although inspiring, can unduly limit research. But a more modest understanding of evolutionary epistemology would emphasise on the two following and more fundamental projects:

The first such project is a naturalisation of epistemology as passing through population thinking: population thinking is a great step forward in the naturalisation of the study of culture — and thus for the study of scientific evolution — as it is a theoretical framework that either does without macro-social entities, or that explain these entities (such as a scientific theory) using natural entities only in the explanans (mental and public representations as material objects). So, population thinking requires specifying which natural entities constitute cultural phenomena, and the processes through which these entities are distributed in human communities and their habitat. The naturalism involved here is concerned with ontology: one must attempt to explain what macro-social entities (scientific theories, states, institutions, etc.) refer to in terms of natural, or material, entities only. A more radical understanding of this project would be to attempt to definitively eliminate non-natural entities in any scientific explanation, but this reductive project is far from being achievable in our current state of knowledge, and is probably not even desirable (chap. 9). So we have this 'modest reductionist' programme as part of evolutionary epistemology.

The second project, which still belongs to the naturalisation of epistemology, consists in spanning the whole range of phenomena out of which knowledge arises, independently of their disciplinary belonging. Recall Campbell's definition of evolutionary epistemology as descriptive epistemology "taking cognizance of and compatible with man's status as a product of biological and social evolution." In effect, this means that evolutionary epistemology studies: (1) biological evolution, as the cause of the existence and nature of the human cognitive apparatus, (2) cognitive psychology, as the description of the processes through which mental representations are constructed, and (3) history, as the description of the particular chains of events that eventually constitute scientific evolution. This project is naturalistic because it aims at showing the connexions between natural sciences, such as biology, and the social sciences. If one renounces to a pan-disciplinary Darwinism, then one remains with the observation that the construction of knowledge involves several layer of constructive processes that may have their own properties. The naturalistic best part of evolutionary epistemology is an integrated constructivist research programme. There are layers of processes constructing elements for the next layer of processes: biological evolution constructs biological cognitive apparatus that construct, when interacting with the environment, representations, which are elements out of which scientific knowledge is made.

Constructivism is a term that cover different theoretical stances but essentially denotes the position that what exists, exists in virtue of a "history of building" (Hacking, 1999). So an integrated constructivist model accounting for the evolution of scientific knowledge aims at investigating the construction of scientific knowledge from its evolutionary stages of natural history, through the cognitive stages of scientific cognition and up to the social stages of cultural construction and maintenance. This does not imply that their need to be a temporal hierarchy from cognitive to social stages: cognitive stages form indeed the building blocks for social stages, but social stages also inform cognitive stages. In particular, enculturation is made of social events that will partly determine later cognitive processes. It is also hypothesised that the evolution of our cognitive abilities has been determined by the social context in which ancestors of the human species have evolved. Constructivism asserts that the development of knowledge is not a process of *extraction* of reliable information from the world; rather, knowledge is acquired through processes that causally involve and are partially determined by cognitive events, in the case of cognitive constructivism, and social events, in the case of social constructivism.

The construction of mental mechanism or cognitive abilities obeys, to the extent that they are innately determined, to the principles of evolutionary biology. Evolutionary epistemology includes evolutionary psychology. There is therefore a process – genetic variation and natural selection – that allows the construction of biological organs, including the human brain. Darwinism is thus opposed to creationism in much the same way as constructionist theories in sociology oppose rational reconstruction or direct realism: neither current life forms nor current knowledge is given, both exist because of some causal processes. In empirical psychology, constructivist theories go back to Piaget's studies on the cognitive development of children. Intelligence, he said, "organizes the world by organizing itself" (Piaget, 1937, p. 311). While it is assumed, in the light of the contemporary findings in psychology mentioned above, that Piaget largely underestimated the cognitive endowment of little children and the importance of innate capabilities, there remains the constructionist claim that intelligence organizes the world. The claim is to be understood as Kantian, asserting, in other words, that one's phenomenally given and cognized world is being shaped by the organisational principles of one's cognitive capacities. Within cognitive studies of science, Nersessian talks of the "constructive practices" of scientists thinking out their research articles (1995, p. 205) and puts a major question of cognitive studies of science in the following terms "How are genuinely novel scientific representation *created*, given that their *construction* must begin with existing representations?" (2002b, p. 133, my italics). Constructivism is therefore a term that can apply to most research in cognitive studies of science, as is made manifest by Giere provokingly titling his response to criticism from a social constructivist, Pickering, "the cognitive construction of scientific knowledge" (1992), thus meaning that naturalist investigation of the processes, the construction, leading to scientific knowledge were not the monopoly of social scientists. In sociology, constructivism has been much criticised as it includes numerous schools of thoughts with views ranging from ultra-relativistic to simply denying that some essential attributes ground some given social status. In the sociological tradition of the Strong Programme, scientific knowledge is socially constructed because it implicates social norms and conventions.

In brief: the processes that lead to biological constructs, cognitive constructs and cultural constructs are not necessarily of the same kind. The biological stages are indeed characterised by blind variation and selective retention, but the cognitive stages are achieved through the functioning of domain specific abilities, including heuristics, naïve theories and metarepresentational abilities, finally, the cultural stages involve, of course, social interactions allowing mental and public representations to stabilise within the population of scientists, through processes such as education, feedback loops, etc. In the next section, I review and specify the constructive mechanisms of scientific knowledge production that I have been describing in the thesis.

12.2 The scientists' mind as being massively modular

12.2.1 Science and the modular mind: why it matters

The first constructive process taking place in the history of knowledge building is the biological evolution of cognitive abilities. Evolutionary psychology holds that the biology of the brain is a product, as any organ of a living organism, of evolutionary history. As a consequence, one can fruitfully understand the brain as having functions for which it has been selected, i.e., functions that increase the chances of survival and reproduction of the organism endowed with the brain in the environment where it has evolved. The overall function of the brain is to process information in such a way that it causes its owner organism to behave adaptively in his environment. Because the brain has been selected by evolution, one can assume that it is successful in this task, i.e., that it implements ecologically rational cognitive processes.

A second important point brought up by evolutionary psychology regards the plausible architecture of the mind. Experimental psychology, esp. developmental psychology, shows that the human mind is endowed with domain specific cognitive abilities with strong innate underpinning. This is the nativist thesis I have been defending in chap. 5. In an evolutionary psychological perspective, these domain specific abilities are best thought of as biological devices with cognitive functions, or evolved cognitive abilities, or "modules." Saying that psychological capacities are modules is, at a minimum, stating that their cognitive implementations are autonomous mental mechanisms (as opposed to merely "functionally individuated cognitive mechanism"—a trivialisation of the notion of module that Fodor (2000, p. 56) rightfully criticises). Evolutionary psychology requires modules to have had an evolutionary history, with the most probable such history being the selection of cognitive devices because of their fulfilling cognitive functions that increase the inclusive fitness of the organism. Thus, Campbell's vicarious epistemic mechanisms are, in up to date terminology, mental modules. Several characterisations may be attributed to modules, such as informational encapsulation, mandatoriness, innate-specification, or implemented in dedicated brain devices or neural systems. These characterisations are debated in the psychological literature and can be more or less plausible and compatible with the ideas that cognitive processes are modular at the conceptual level (i.e. it is not only perception and other "peripheral systems" that are modular) and that the architecture of the mind is "massively modular." The massive modularity hypothesis asserts that the mind consist almost entirely of modular systems. In particular, domain-specific abilities are subserved by modules, almost all of the processes that generate beliefs and decisions are modular in nature, and there is no such thing as general learning subserved by some allpurpose cognitive device. The reason for denying that the human mind includes a domain-general and all purpose cognitive device is that the advent of such a device during the evolution of the brain is very implausible. While evolutionary psychologists have supported their claims about the existence of modular constituents of the mind by reconstructing the phylogenetic history of these constituents, there seems to be no such support in favour of a general purpose cognitive ability. On the contrary, any plausible natural history of the mind requires it to evolve through relatively small numerous changes, which, given the increase in cognitive abilities and brain size in the natural history of our species, seemed to be largely incremental. This introduces the notion modularity as follow:

The Cartesian view of a seamless whole makes it hard to see how such a whole could have come into being, except perhaps by an act of divine creation. By recognizing the modularity of mind, however, it is possible to see how human mentality might be explained by the gradual accretion of numerous special function pieces of mind. (Cummins and Allen, 1998, p. 3)

A similar argument is already present in Simon's idea of near decomposability: evolution could have produced a complex system only by adjusting components piecemeal, which then compose into wholes (1996). Complex devices can be constructed but only in a piecemeal manner. Complex organs and devices, such as the brain, come into existence through a process of incremental complexity. Concerning the phylogeny of our cognitive apparatus, an attempt to retrace its natural history could go along the following sketchy line. Presumably, the evolution of adaptive behaviour has for ancestor mechanical responses to variation in the environment, such as sunflowers turning their heads toward the sun. It begins with simple neuronal devices and eventually issues reflexes stimulus-response, such as the frog automatically throwing its tong at small black flying object. It continues with an enrichment of the treatment of the input, which leads to a finer parsing and analysis of the environment, and a larger range of behavioural responses. It includes an increase in the computational steps between stimulus and responses, allowing much information, such as memories of past events or planning of future events, to sneak in and inform the responses. The evolution of an all purpose cognitive device, by contrast, would require an unlikely variation upon the cognitive apparatus of the species from which we are issued. Moreover, a domain general ability would require much more computation, and thus time and energy, for solving tasks for survival than any task-specific module. This is because while a task specific cognitive device can promptly call on an already made small database, simple heuristics or set of procedures for solving the problem at hand, a domain general ability would need to first perform several complex and lengthy computations such as an analysis of the problem and ensuing goals, and an evaluation of the possible behaviours. In the face of evolutionary selection, organism endowed with cognitive devices that lead to adaptive behaviour quickly and with little consumption of energy is much more advantageous than domain general cognition. So if domain general ability were to appear, it is unlikely that it would replace the quick and simple, but yet adapted, cognitive processes of task-specific abilities.

Arguments in favour of the existence of central-modules, as domains specific abilities with innate bases, are found mainly in neuropsychology, which has discovered selective impairment of given abilities; in brain imaging, which attempts to localise the physical implementation of cognitive processes; in developmental psychology, which has discovered precocious abilities in a number of domains, such as naïve physics and naïve psychology. Some results also come from cross-cultural psychology, which has advanced evidence of the existence of cross cultural modular abilities, such as naive biology. On top of these experimental results, evolutionary psychology brings up a further theoretical argument in favour of the massive modularity hypothesis. It is a constructivist argument, as it compares the probabilities that evolution 'constructed' a massive modular mind rather than a non-modular mind. In this perspective, mental modules are considered as evolved psychological building block — with most blocks having an adaptative function which explains why they have been selected. The argument is not a knock down argument against a view of the mind that assert the existence of a general purpose cognitive ability, but the challenge it raised — can you furnish a plausible history of the cognitive abilities whose existence you hypothesise?— has not been satisfactorily met by the proponents of such a view.

Evolutionary psychology and the massive modularity hypothesis lead to view the architecture of the mind were evolved abilities form a relatively stable basis for cognition. Changes in the cognitive organisation of abilities are thought to be fewer and less drastic than in some competing theories of the structure of the mind (Karmiloff-Smith, 1992, e.g.). One reason is that in the absence of domain general cognitive devices, changes in the organisation of cognition are left to domain specific abilities. Presumably, these abilities can effectuate changes only in their own area: for instance, the language learning module may produce language abilities, but it cannot have a direct impact on other modular abilities. Another reason is that the cognitive abilities that have been selected during evolution are tied to the genetic endowment, which does not change during individual history. This, of course, is not to deny the role of development and the environment in the construction of cognitive abilities: for instance the natural languages that one learns depend on one's environment. Yet, cognitive traits that are adaptations, for which the genotype has been selected, must be relatively robust so that their proper functioning (the reason why they have been selected) is not contingent on small variations in the environment. I will call cognitive plasticity the view that new abilities, as functional cognitive entities with some specific biological basis, can easily be formed, i.e. the view that learning can importantly change the architecture of the mind.

It is at first hand difficult to picture the scientist's mind as being massively modular. This is because scientific cognition is creative in ways that seems prevented by the constraints on cognition set by the massive modularity hypothesis. Thus, Fodor (1983, 2000) argues that central cognition, in particular the processes issuing in belief formation, are not modular. The arguments in *The modularity of thought* (1983) appeal to scientific cognition as the archetypical cognitive performance, which shows that belief formation relies on cognitive processes that can draw on *any* information held in the mind. Scientists, or so it seems to Fodor, have an unrestricted access to their stored information, which couldn't be if the human mind was massively modular. Fodor's pessimistic conclusion is that cognitive psychology can never account for belief formation. This is because the computational theory of mind is the only remotely plausible account of how the mind works, yet computational psychology has no means to bound the set of beliefs to consider for hypothesis formation, other than by assuming that the mind is massively modular. Showing that scientific cognition can be implemented by massively modular minds has therefore a high stake: it would be a key denial of Fodor's pessimistic views on the future cognitive psychology. There is therefore a methodological argument for the massive modularity hypothesis: in spite of the difficulties it comes with, as those forcefully pointed out by Fodor, the massive modularity hypothesis remains the best available research programme — cognitive processes can be analysed only if there is some computational tractability; tractability necessitate informational encapsulation; encapsulation is the main criterion for modularity; so either the mind is massively modular, or cognition cannot be analysed.

Scientific cognition is an extreme instance of cognition — it represents only a minute portion of human cognition, both across human history and in our present days where science is highly praised. Yet, it has been central in cognitive psychology as an inspiring source, especially as the archetypical example of human reasoning. The last step is misguiding when it is the folk conception of scientific cognition that is taken as archetypical, rather than what scientific cognition really is. A first criticism along this line can be found in Hutchins' interpretation of the history of cognitive psychology (1995). He argues that the mistaken internalism of traditional cognitive psychology originate in Turing's folk analysis of how one perform mathematical reasoning. Turing and his internalist followers, Hutchins argues, missed the fact that mathematical reasoning relies on distributed cognition. Can't the argument be extended to Fodor? Isn't he liable to the same mistake of taking his folk understanding of scientific cognition as paradigmatic?

12.2.2 HOW MASSIVE MODULAR MINDS CAN BE FLEXIBLE: PROPOSALS FROM A COGNITIVE SCIENCE OF SCIENCE PERSPECTIVE

Cognitive flexibility is defined as the ability to adapt cognitive processing strategies to face new and unexpected conditions in the environment. It involves learning how to deal with new types of problems by implementing new computations. These learning abilities seem not to be attainable with massively modular minds — which are composed of task specific cognitive devices. The massive modularity hypothesis also imposes important constraints on the architecture of the mind and on the consequent flow of information: an input is processed by the modules to which it meets the input conditions, which produces an output acting as an input for further modules, depending on the architecture of the mind, till the processing come to a halt. Modules the communication between modules is relatively limited, and strongly constrained by the relatively rigid cognitive architecture. How can we account, with this hypothesis, of the known flexibility, diversity, malleability and creativity of human behaviour? How can we account for the human ability to integrate information from different domain? It is a challenge that proponents of the massive modularity hypothesis have taken seriously. Sperber (2002, 2005) argues that flexibility and context-sensitivity are attained, at the psychological level, because most modules are learning modules. Learning can happen not only through enrichment of modules' databases but also through the fixation of parameters determining the domains of modules. Nested modularity, maturation of cognitive abilities through interaction with the environment, enrichment, and many other processes endow modular minds with much more flexibility and adaptive potential than might initially be thought. Development, according to Sperber, also includes learning that is reflected on modular architecture: learning modules produce dedicated modular subsystems for acquired capabilities.

Context sensitivity

In order to account for context sensitivity, Sperber, further argues that modules do not process inputs in a mandatory way. Mandatoriness is one of Fodor's characteristic of modules. It implies that once an input meets the input conditions of a module, the module is automatically triggered and run its full course. Sperber argues on the contrary that input are processed by a module only if they meet its input condition, but also only when they are sufficiently relevant. In other words, modules ignore plenty of input meeting their input condition because processing the input does not issue enough new information. Carruthers (2003) argues for a 'moderately massive modularity' where the language module is given a special role as serving as the medium of inter-modular integration and conscious thinking.

Without denying the role of the above principles of flexibility, context-sensitivity and integration, I would like to emphasise the role of meta-representations in generating new integrated knowledge, and as eventually sustaining conceptual change in science. The flexibility of the human mind, indeed, is paradigmatically exemplified with conceptual change in science, where some previously held beliefs are abandoned and replaced by new beliefs incommensurable with them. In particular, conceptual changes in science have rendered some of the content of science at odd with intuitive beliefs. How can we have come to think, and be now so convinced, that the earth is moving around the sun while the contrary belief naturally imposes itself upon us? While knowledge enrichment can be thought of as the addition of new data to previously existing databases, conceptual change and abandonment of previously believed theories requires, on the part of the scientists, a new attitude towards the stimuli of the newly theorised domain. What are the cognitive processes accounting for these new attitudes? Conceptual change is a key problem in science studies and an account of it needs to include the events in people's minds that make these conceptual changes possible.

The existence of conceptual change raises two questions for cognitive psychologists: first, what are the cognitive processes that make conceptual change possible? Much work has been done in cognitive studies of science on this topic. Most notably, Nersessian (1992a) has analysed the role of physical analogy, the construction of thought experiments and limiting case analyses. Carey has also pointed out the role of mappings across cognitive domains for the creation of new domains (e.g. Carey 1985; Careyand and Spelke 1994. There is general agreement that conceptual change involves metarepresentational abilities; the debated point is on the necessary development of these abilities and their complexities for conceptual change to be possible (see Carey and Johnson 2000). The second question is: What are the cognitive processes that are implemented once conceptual change is achieved? I will focus on the second question.

What is at stake, for cognitive studies of science, is whether the innate abilities mentioned above (naïve theories, innate heuristics) are put to work in scientific cognition, or whether scientific cognition relies on other abilities that develop during ontogeny, and in particular through education. If we are in the latter case, then cognitive studies of science would have to concentrate on the role of the acquired abilities at work in scientific cognition rather than on the innate abilities designed by evolution. Evolutionary psychology would then be much less relevant (or less directly relevant) to science studies and the task would be to discover, with other means, the *developed* cognitive abilities sustaining scientific cognition.

Massive modularity hypothesis, meta-representations, and scientific thinking (scientific cognition as presented in the second part of the thesis and put in a massive modularist perspective)

The picture of scientific cognition I have developed in the second part of the thesis ('psychology and the history of science') is as follow:

The scientist's mind is made up of modules that implement ecologically rational cognitive processes, so the existence of scientific cognition raises the following questions: How can a species that evolved as a hunter-gatherer species do science? How can we obtain scientific rationality out of people's ecological rationality? How can we have gone beyond biologically implemented cognitive heuristics, innate naïve theories, or psychologically interpreted Kantian categories to obtain our scientific understanding of the world? My first attempt to answer these questions consists in answering a more specific one: How and why are cognitive modules put to work on scientific problems?

Scientific cognition, I argued, heavily relies on the ability to meta-represent our own representations, and thus to think reflectively. Meta-representational ability allows for the processing, using and producing of representations of representations. The ability may be implemented by one or more cognitive modules. Some meta-representational modules, indeed, have an already studied evolutionary history and satisfy the requirements of evolutionary plausibility. Presumably, meta-representational abilities appear with the ability to represent the representations that others may hold – their mental state. This ability, called Theory Of Mind (TOM), is adaptive by allowing Machiavellian intelligence, the ability to manipulate others' behaviour, and is certainly at the basis of human social life, including linguistic communication.

The relevant consequence of meta-representational ability (or abilities) is that the product of modules can be re-thought. In other words, mental representations can be taken as input of metacognitive abilities so as to provide meta-representations that will determine the attitude one will hold with regard to the input representation. For instance, one can think that the input representation X provides a true or a false representation of the world through having the metarepresentations 'It is true that X' or 'It is false that X'. Metarepresentations can also express semantic relations among representations (e.g. X contradicts Y) and evidence for beliefs (e.g. A justifies my belief that B) (Sperber, 1996). More generally, metarepresentational abilities allow for the interpretation of representational output of previous (modular) heuristics and naïve theories; these representations can be reflected upon and given some further meaning through the embedding of representations. The most obvious case is when sounds uttered by some speaker are interpreted as conveying what the speaker intends to communicate (Sperber & Wilson, 1986), but interpretation is also at work when our intuitions are taken to reveal something about the world rather than directly leading to (adaptive) behaviour. This happens, for instance, when perceptive representations get embedded within a framework theory; then, the perceptive representation is metarepresented as a manifestation or consequence of some state of the matter or laws of nature. We look at the light of a bulb as being a consequence of moving electrons, for instance. Cognitive studies of science have not ignored the pervasiveness of metarepresentations in science. Scientific practice, says Nancy Nersessian, "often involves extensive meta-cognitive reflections of scientists as they have evaluated, refined and extended representational, reasoning and communicative practices" (Nersessian, 2002a, p. 135). Deana Kuhn has also pointed out the metacognitive skills at work in scientific thinking. These include not only meta-strategic competence, but also the ability "to reflect on one's own theories as objects of cognition to an extent sufficient to recognize they could be wrong" (Kuhn, 1996, p. 275). Metacognition and other more basic metarepresentative abilities are thus central to scientific thinking. Most interestingly for our present purpose, they also bridge the gap between lower cognitive abilities processing the input from our sense organs and other hardwired heuristics or naïve theories, and the abstract and consciously controlled thinking practices of science. In particular, problem representation consists in bringing a set of representations and previous knowledge or ideas to bear on the understanding, or interpretation, of incoming 'naïve' or intuitive representations. Problem representation allows cueing heuristics in the search of solutions. Gorman (2000) illustrates this point with Kepler's mental model of the solar system and the application of heuristics as designed and implemented in the discovery program BACON 1 of Herbert Simon and his colleagues. Kepler's particular problem representation, he explains, was necessary for the heuristics to apply and be useful. In general, the interpretation of naïve or intuitive representations makes possible directing them further towards other heuristics, naïve theories or any modular processes. For instance, our interpreting of electric phenomena as a consequence of the movement of electrons activates our naïve physics theories. In those cases, metarepresentations act as routers of representations towards the right module. The routings therefore make use of ecological rationality for the development of our understanding of the world and the construction of a scientific rationality that is oriented towards truth rather than increase of inclusive fitness. This development of scientific cognition is a cultural achievement because evalutative, interpretive and routing meta-representations have been developed with scientific theories and practices during the historical evolution of science. Thus, the 'right' that qualifies the choice of modular processes and the 'scientific' or 'rational' that qualify the thoughts has now to do with the normative aspects of scientific traditions and paradigms. Problem solving using heuristics, of course, is both learned by humans and biologically given. One way the learning can happen is by using already existing heuristics for solving problems that the heuristics were not initially designed to apply to. This use of heuristics for ends they were not originally created for is described as 'exaptation' by Wimsatt (2000). He notes that "evolution, human engineering, science, and culture all systematically reuse constructs in new contexts that drive their elaboration in new directions". I suggest that scientific thinking is well characterised as a systematic exploitation of human cognitive abilities by constructing, via metarepresentations, exaptative heuristics and intuitions. Specifying the role of meta-representation in scientific cognition and which evolved domain specific abilities are put to work in given scientific contexts provide a psychologically informed basis to the general picture of scientific thinking provided by (Barnes et al., 1996, p. 127):

> The machinery involved in the perception and recognition of things hums along undisturbed much of the time. For the individuals in a given culture it usually hums along in unison; indeed it has so to do for the culture to exist. The fact of its existence depends upon a certain blind conformity in perception, understanding and judgement, in initial responses to things. But the machinery of perception and recognition is nonetheless subordinate to reflexion and calculation. The basis of sociability, and thus of humanity, lies in our shared tendencies to automatism, but its actual achievement lies in the calculative exploitation of these tendencies.

An important gap in science studies is the study of the role of our primary intuitions in scientific knowledge. Social studies accord little importance to these cognitive events that are intuitions, while cognitive studies are much more focused on higher reasoning practices (induction, abduction, analogical reasoning, thought experiment, etc.). The continuity thesis, which asserts that scientific cognition is of the same nature as lay cognition, has raised important debates that could bear on the distinction and relation between reflexive and intuitive thinking, between metarepresented knowledge and direct output of non-metarepresentational modules (see Sperber, 1997, for the distinction between intuitive and reflective beliefs). However, the empirical stake of the debate has not focussed so much on the use of common sense in scientific cognition (with the exception of Atran 1990) as on whether the higher reasoning practices of scientists are used by laymen and children. Concerning the normative rational practices, such as the use of deductive logic, psychologists have found that laymen mostly do not follow them, and thus do not answer the normative criteria. On the other hand, most theories in developmental psychology have asserted that children do think in similar ways to scientists, including hypothesis-testing, theory formation that allows them to develop theories that are incommensurable with the theories they replace, and general processes of belief formation leading to the 'scientist as child' metaphor (Gopnik, 1996). That norms of reasoning may not be followed by lay people comes as no surprise from the perspective of ecological rationality; and it is also unsurprising that creative thinking in children and adult scientists relies on the same cognitive processes and abilities; creative thinking is not based on a cognitive ability that develops only when doing science.

What of the role of naïve theories, biologically implemented heuristics, and 'lower' cognitive processes with percepts as output? These are innate endowments that provide our unconscious and non-reflexive thinking and guide most of our actions; they are pervasive in day-to-day cognition, but their content may be inconsistent with contemporary scientific theories. People do not reason with quantum mechanics for grasping things and we do not normally think of ourselves as moving in a Riemann space. Scientific knowledge is not embodied in the innate endowment of the human mind. Does that mean that the study of this endowment is irrelevant for the study of scientific cognition? In other words, is this endowment fully bypassed and of no consequence in scientific cognition? My answer was a radical no, because scientific cognition always draws on the inferential resources of modular abilities — but what if these modular abilities change with scientific enculturation? It is a rather peculiar alchemy, one must admit, to educate someone who is born for survival and reproduction so as to make a scientist out of her. Studying the properties of the scientific mind implies studying first the brain as an evolved biological organ, and then the transformation that scientific enculturation brings about. Scientific enculturation is among the key constructive processes knowledge making: it is the life time constructive process of key cognitive elements — scientists' brains — in the production of scientific knowledge.

Conceptual change without modularisation

I argue that the cognitive processes allowing conceptual change have little effect on the architecture of the mind; all that is needed is enrichment of meta-representational knowledge. The same intuitions and abilities sustain pre-conceptual change and post-conceptual change cognition. I defend a strong continuity thesis, which asserts that the lay man and the scientist have much the same cognitive abilities organised in a similar way. Scientific knowledge does not necessarily induce change in the cognitive architecture. In particular, scientific knowledge, and scientific conceptual change, does not inactivate modules, even when the knowledge contradicts the output of the module. It does not induce module replacement. So the scientists' mind has the same modular architecture as the lay man, in spite of his different understanding of the world.

By contrast, Carey and Gopnik defend a weak continuity thesis which asserts that only the discovery processes need be identical in child cognition and scientific cognition. They hypothesise that conceptual change in science is based on isomorphic changes in people's mind. For Careyand and Spelke (1994), conceptual change in scientific development necessitate other processes than those offered by meta-representational abilities. "Reflection by itself", they say, "will not produce conceptual change" (p. 180). More radical proposals have been advanced by Churchland, Gopnik and Karmiloff-Smith, which bet heavily on the cognitive plasticity of the human mind and/or on domain general processes that cause modularisation. These theories are hardly compatible with current theories in evolutionary psychology. Carey's framework, however, is the most compatible with the approach defended here. Carey distinguishes core theories from intuitive theories: core theories are those theories that are innate endowment and which account for the behaviour of infants, while intuitive theories are constructed during cognitive development. Examples of core theories are naïve physics, naïve psychology, and naïve quantitative reasoning. Examples of intuitive theories are number cognition, after, among other things, the integration of the concepts of zero and infinity and the construction of mappings between numbers and geometry. Children also develop, Carey argues, an intuitive theory of biology, which arises after conceptual change in the concept of living things. A third example of intuitive theory is provided by conceptual change in the years 4 to 12 in the interrelated concepts of matter, weight and density (see Careyand and Spelke 1994, pp. 184–194).

For Carey, core theories are modules in a sense akin to the one already used in this paper, but Carey further takes intuitive theories to be modules, thus rejecting the criterion of innateness as a necessary property of cognitive modules (1995, p. 274). The problem is that Carey's consequential identification criteria for modularity are too weak. Fodor (2000), as already mentioned, warns us against the temptation to take modules as "functionally individuated cognitive mechanism," as these means of individuation would not allow to tackle the problem of computational tractability using the concept of modules. More importantly, I think, functional individuation leaves untouched the problem of the biological implementation of cognition. The problem is not merely terminological and there are some reasons to insist that modules be defined as cognitive organs, and thus answer some criterion of biological implementation, as for instance, through which process has the modules been biologically constructed. There is the methodological choice between semantic criteria of identification (Carey: identifying a theory on the basis of which people explain the phenomena pertaining to its domain) versus realist-existential assertions about the structure of the mind (Sperber-Atran: identifying cognitive organs). Semantic analysis of the cognitive processes is certainly the best analysis, if not the only possible one, for cognitive psychology. But the integration of cognitive psychology with biology – from either brain imaging or evolutionary psychology — imposes and allows stronger, existential, claims for 'modules'. The integration is desirable not only for the reduction of semantic properties to biological ones – a naturalistic programme of its own — but also because the semantic functioning of the mind is likely to be highly constrained by its physical implementation. The embodiment of the mind is likely to have some consequence on its functioning, so one is not only interested in a detailed account of what the mind does (a semantic-functional account), but also in how it actually does it (a realist account). In particular, there is no a priori reason that later cognitive achievements be implemented in the same way as modular abilities or have the same epistemic properties. Cognitive processes implemented through genuinely modular abilities are even probably very different from cognitive processes implemented through intuitive theories.

An analogy between physiology and the architecture of the mind may suggest an answer for the problem of the implementation of intuitive theories. Notice, indeed, that we can use our hands and liver in ways that are certainly not the function they have been selected for. Organs can enlarge their actual functioning beyond the limits of their evolved designed function. For instance, we can use our hands to play the piano, while they certainly have evolved for grasping, and we can use our liver for the digestion of Champagne, while it more probably has evolved for digesting the food consumed by hunter-gatherers. The hypothesis is therefore that humans can use their cognitive modules in novel ways, for which they were not designed by evolution. This, in turn, can lead to conceptual change and the development of new intuitive theories. Atran and Sperber's work (Atran, 1990, 1998; Sperber, 1996a, 1997a) provides an account of theory change along these lines. Implementations of new theories, they assert, do not replace modular abilities, but on the contrary continue to rely on them. Atran uses neo-Darwinian theory to illustrate theory implementation without replacement. Neo-Darwinian theory has a notion of species as sets of organisms that live in the same ecological niche and that can interbreed. This notion is incommensurable with the naïve notion of species, which is essence-based and associated with a favoured rank within the folk taxonomy. However, Atran argues, the adoption of neo-Darwinism does not cause the elimination of naïve pre-theoretical intuitions. What happens, rather, is that naïve thinking still provides the basic intuitions and percepts upon which scientists reflect so as to interpret them within a neo-Darwinian framework. So the ecologist doing fieldwork still perceives animals as entities at the generic species level and with essences as intrinsic teleological causes. But in his university office, the same ecologist will interpret his data thus gathered through his basic cognitive abilities, especially naïve biology, in the light of the most recent scientific theories. Naïve biology is a cognitive organ that presumably evolved as an adaptive skill for the hunter-gatherer (we can also suppose that some kinds of naïve biology are present in other animals' cognitions); it is nonetheless put at work to do science, a function it did not evolve for. Scientific reflection upon the output of the naïve biology module bestows a theory that is inconsistent or incommensurable with naïve biology. The new theory does not emerge through the transformation of the module, which continues to provide the same intuitions and percepts; it emerges due to a reflective attitude upon the module's output. The cognitive processes sustaining the theory therefore lie in the functioning of the naïve biology module, together with some sets of meta-representations which provide the context for the interpretation of the outputs of the module. Scientific enculturation need not generate new cognitive structures. It is, on the contrary, implemented through enrichment only, consisting of beliefs that will constraint future interpretations and reflections upon our primary intuitions.

For Carey, some core theories might be overthrown and replaced by intuitive theories: this is conceptual change. For Sperber and Atran, core knowledge (or naïve theories or modular abilities) are biological endowments that do not disappear with cognitive development, even when knowledge that is inconsistent with core knowledge is elaborated. Beliefs obviously change, but this does not alter or transform the architecture of the mind, which consists of an arrangement of modular abilities constraining information flows. Thus, change of beliefs and the evolution of knowledge do not create new intuitions or perceptive abilities (neither ontogenetically nor historically); new beliefs, cultural and historical variations, always rely on the same basis of intuitions: the output of biologically realised cognitive devices. This hypothesis is corroborated by the fact that scientists act and think in everyday life exactly as laypeople. The expert in quantum mechanics continues to see a cup as a cup, rather than as a complex of interacting elementary particles; the biologist, as Atran points out, continues to see the tree as a tree, even if this category has no scientific counterpart; and the psychologist continues to understand people as intentional agents, even when he adopts the most radical behaviourist theories. In a sense, contemporary science is highly unintuitive, but while beliefs vary greatly, intuitions and perceptions varies comparatively little.

Another difference between Carey and Sperber-Atran lies on the biological basis of mental theories that develop during ontogeny. Because Carey uses a semantic criterion for distinguishing abilities, she is not able to distinguish between abilities that reflect the working of a (biological) cognitive device and abilities that cut across and use several cognitive devices. A semantic criterion is not sufficient for the circumscription of cognitive domain of (biologically realised) modules. Carey consequently postulates the existence of mental devices that develop during ontogeny: the intuitive theories. These intuitive theories have a status in between modular innate abilities and scientific theories. They are mental devices as innate modules, but they are developed in the same way as scientific theories. In their argumentation against Sperber, Carey and Spelke present intuitive theories as the necessary mental ground of scientific theories. The counter-argument that I have presented, however, consists in showing that the mental ground of scientific theories need not be a mental cognitive device of its own that somewhat mirrors the content of scientific theories. Such a view seems to stem from a persistent simplification of the constitution of scientific theories, which are reduced to sets of beliefs and its ensuing reasoning abilities, i.e., a semantic characterisation of scientific theories. With this simplistic view, the development of science and conceptual change is indeed in need of its mental counterparts: the same theories and change put within the mind of the scientists. Consideration of the problem of the physical implementation of scientific theories, however, raises new problems and shows the limits of Carey's purely semantic analyses. At the level of the brain, I have argued that scientific cognition is implemented by modular primary abilities together with reflection – including semantic evaluation – upon their output. Likewise, Erana and Martinez (2004) have argued against a semantic reductive view of scientific theories, pointing out the complex of mental, cultural and artifactual interacting components of scientific theories. Taking into account the physical implementation of the scientific theories outside the brains of the scientists similarly allows Erana and Martinez to argue against Carey's and in favour of Sperber-Atran's view of cognition.

Let me clarify my criticism: I fully agree with Carey and Spelke that there must be some mental implementation of the semantic content of theories. I have no argument against the idea of intuitive theories, insofar as they describe semantic properties of people's cognition. But I have invoked the importance of specifying their physical implementation and rejected the hypothesis that scientific theories are implemented by their own cognitive device. I have hinted at an account of an implementation of non-innate knowledge and theories by calling on the working of biologically realised modular abilities, including meta-representational abilities. I have defended the theory that asserts that the cognitive architecture of the mind is relatively stable and varies little with scientific education. At first glance, this may appear to contradict our knowledge that beliefs, scientific beliefs included, greatly vary in space and time. But the historicity of science is not a counter-argument to the thesis that there are strong innate constraints on mental cognition. On the contrary, it is the innate mind that provides the dynamics of scientific development. I have thus sketched a view of the mind where conceptual change is implemented through the working of pre-conceptual change cognitive devices and the processing action of meta-representations. The latter can feed in modules with new representations, thus exploiting the module processes for further inferences – this is what happens, for instance, when the light of a bulb is understood as the manifestation of the movement of very small objects (electrons). Meta-representations can also distinguish among illusory and revealing intuitions through giving them a semantic status. They provide new meaning to these intuitions by embedding them in acquired knowledge. Scientific cognition is then described as a culturally informed reflection, allowed by meta-representational abilities, upon the output of preliminary modules delivering intuitive beliefs. A main counter-argument to the massive modularity hypothesis consists in asserting that the specific adaptiveness of the human species comes from its cognitive flexibility and inventiveness leading to adaptiveness to new situations. My argument has been that much of flexibility is due to our metarepresentational abilities. The proposal is akin to Carruther's proposal, attributing the origin of flexibility to language, but locates the source of flexibility at the more basic level of metarepresentation - which need not include linguistic items.

Situatedness and environmental embodiment of cognition as means of flexibility (scientific cognition as presented in the third part of the thesis)

It is noteworthy that Fodor's examples of cognitive achievements, which are presented as *not* being possible output of massively modular minds, are cultural achievement. In *The modularity of Mind*, the example is scientific knowledge; in *The mind doesn't work that way*, an example is a cooking recipe, about which Fodor asks how we came to think about putting such ingredients together. It may happen that Fodor is absolutely right and that science and cooking recipes cannot result from the computations of a massive modular mind. But his conclusion — that the human mind is not massively modular — is not warranted for all that, since the cognitive achievements result not from one single isolated mind, but from people with each others, with the world, and through historical developments. Fodor's arguments against the massive modularity hypothesis hold only if one can show that the cognitive achievements he appeals to cannot result from cognitive processes that span several massively modular minds interacting with their environment. However, it is the contrary that is true: historicity and the environmental embodiment of cognition bring the necessary flexibility for the consequent cognitive processes to issue Fodor's example of knowledge. The second part of the thesis has shown that cultural representations, as present in the surrounding material environment of the scientists, as having participated to the development of the scientits' minds (if only by increasing knowledge), and as known to be present in the scientist colleagues' minds strongly determine scientific cognition. In particular, cultural representations form interpretive traditions with which scientists interpret and produce further scientific data. What happens, therefore, is that the cultural environment change, and that this causes human cognition to change. Fodor points out that in scientific reasoning, anything can be made relevant to one's topic, and any proposition can enter one's reasoning. This, he maintains, renders the cognitive processes involved untraceable. Spranzi's (2004) case study is an example of such reasoning where an analogy is drawn between two distinct phenomena: Galileo interprets the black marks on the moon as similar to the shadows thrown by mountains on the earth. Now, Spranzi argues, the analogy did not pop up out of the blue – which would have exemplified a mysterious 'Fodorian' (isotropic) cognitive event. She shows, on the contrary, that it was rendered possible through a historical process of bootstrapping. In other words, the cultural context made some ideas and representations available to Galileo, thus framing his cognitive environment (Sperber and Wilson, 1986, § 1.8) and making the analogy possible. We therefore have a case where the determination of scientific thought is shown to be historical and social as well as cognitive. The mystery is solved by realizing that cognition takes place in a cultural environment, which is historically constructed. There is a process of co-evolution of scientific cognition and culture.

In the previous sub-section, I have argued that the changes caused by scientific enculturation can be accounted in a massive modularist framework: meta-representational abilities are such that they enable scientific cognition to be context dependent, and, more precisely, dependent on currently held scientific (and to a controversial extent non-scientific) beliefs. Flexibility of human cognition is accountable with its cultural situatedness. By this, I mean that (1) human cognition is characterised by its numerous interactions with, and its strong reliance on, the environment; (2) scientific cognition makes extensive use of this part of the environment that has been shaped by past people (3) that changes in the cultural environment is a factor of flexibility. Flexibility in human cognition is attained with the loop between the mind's cognitive processes and the changing environment: the mind's processes determine behaviour which changes the environment, which cause change in the psychological processes that determine behaviour, etc. This loop is all the more important when one realises that cognition actually takes place *in* the environment — cognition is distributed. It was the aim of the third part of this thesis to describe how scientific cognition is implemented in both scientists' brains and in their actions and environment. Scientific distribution is distributed because information flows through, and is transformed by, the environment. Some representations are publicly instantiated — taking the form of written publications, computer generated drawings, utterances pronounced during colloquiums or during informal exchange, etc. — and they are transformed by the environment — experimental machineries are aimed at producing public representations, they process information, and scientific practices give an increasing role to cognitive tools. Moreover, the flow of information may be constrained in systematic ways so as to complete specific cognitive tasks. When it is so, cognitive elements, humans or not, are given specific places and roles; elements are organised and form some distributed cognitive system.

An important fact about distributed cognitive systems that are socially implemented is that they are not rigid as massive modular minds are supposed to be. On the contrary, distributed cognitive systems evolve during history, taking in new elements, taking out other elements, and maybe changing their very goals. In the case of science, distributed cognitive systems that output scientific knowledge evolve when the boundaries of disciplines change, when the experimental apparatus is renewed, and when the object of scientific inquiries change (Daston, 2000). Here is, therefore, another source of flexibility: scientific cognition is implemented in distributed cognitive systems that quickly change, they have the plasticity out of which flexibility arises. This plasticity of distributed cognitive system enables quick adaptation to changing goals and environment. In particular, new technologies are exploited, and the architecture of the systems changes in function of the available resources and goals (for instance, contemporary large experiments in atomic physics require numerous researchers dealing with very specific tasks, while traditional theoretical debates require few researchers having similar expertises). This suggests that distributed cognitive systems evolve so as to respond to contextual factors such as the changing means and needs. Once one has acknowledged the plasticity of distributed cognitive system, however, one can question the causes of changes: how is this plasticity used to respond to the context? What determines the architecture of distributed cognitive systems? And are ensuing architectures *optimal* or *adapted* responses to the context?

Distributed cognitive systems are social institutions, and their design and systematicity is determined by the distribution of 'regulative representations.' If the distribution and content of regulative representations change, then so does the distributed cognitive system. The representations that regulate distributed cognitive systems are mainly representations attributing cognitive functions to elements. They take the form of 'element X is trustworthy for performing task t', such as 'Dr. X is a competent expert in ADN structure', 'this computer can check the reducibility of unavoidable sets', 'this telescope actually shows the relief of the earth', etc. Representations of trustworthiness generate deferential behaviour out of which cognition is distributed. Input is taken from trustworthy elements, and the specification of which cognitive element takes its input from which other cognitive element is exactly the specification of a cognitive architecture. Distributed cognitive systems evolve by piecemeal allocations of cognitive function, with decisions taken as opportunities and difficulties arise. Much of the evolution of distributed cognitive systems passes by scientists' evaluation of the cognitive worth of elements. These evaluations apply whether the element is human or not, although the questions and cognitive processes involved may be quite distinct (for instance, evaluating the trustworthiness of people implies assessing their interests and benevolence, which calls on one's theory of mind; evaluating the reliability of cognitive tools, on the other hand, may rely more on tracking the ratio of good versus unsatisfactory performance). The evaluations are cognition about cognition. Meta-cognition would therefore be at the centre of the evolution of distributed cognitive systems. Interestingly, meta-cognition is also at the centre of the evolution of interpretive traditions, as these traditions have been characterised as specifying which domain specific capacities are proper, in the scientific context, for thinking about some given sets of phenomena. Interpretive traditions include meta-representations that evaluate the significance of (or reflect upon) the output of primary intuitive abilities. With these evaluations, the interpretive traditions distribute new, culturally specified, functions to mental, modular, abilities. Eventually, scientific cognition is largely based on meta-representational capacities that manage the exploitation of cognitive processes, be there implemented in the brain or in the environment.

12.3 HOW SCIENTIFIC COGNITION EVOLVES WITH CULTURE

Saying that meta-representational capacities manage the exploitation of mental and environmental cognitive processes is not a homunculus fallacy. The fallacy consists in postulating the existence of a little man inside the head which directs cognition. The temptation, indeed, is to understand the innate endowment of the mind, heuristics and naïve theories, as cognitive tools intelligently used by some mental homunculus, who takes the eventual decisions. The theory is fallacious because, among other things, it explains nothing: once a homunculus is postulated, then the problem is just relegated to explaining the psychology of the homunculus. Although the fallacy may appear obvious, it may be easy to commit it under disguise. Is the meta-representational management of cognitive processes a homunculus in disguise? No. The hypothesis does not commit the homunculus fallacy because one can specify how meta-representations come to manage cognitive processes the way they do. No independent power of decision is given to meta-representational abilities, but their occurrence and content is determined by the history of the individual, and in particular, by her past social interactions. For instance, I have shown how representations attributing cognitive functions to computers in proofs came about through historical developments of the discipline of mathematics and its relation with computer science (chap. 7). Interpretive traditions, as determining meta-representations interpreting scientific data, also have a history that account for their existence — this history involves the interests of the social actors, and it involves their responses to past interpretive traditions or theories, which itself is determined by their biological cognitive endowment.

The above argument is based on the fact that the environmental determination of human cognition is not restricted to furnishing input computed by the brain, which then issue adaptative behaviour. The environment participates to the computation of representations, as they are computed in social cognitive causal chains, which span through people and artefacts. Cognitive processing lasts through time, in cultural communities, in distributed cognitive systems, in socially situated and embodied minds. Fodor worried about the tractability of computations in a mind, and argued that the problem could not be solved for the processes of belief formation: the information to compute for belief formation is open ended, as anything may be relevant. So it seems impossible to track which information shall and will be processed. Fodor was right about tractability, but his conclusion was that there was something wrong with computational psychology, while the problem rather lies in the internalist credo of traditional cognitive psychology. The tractability problem may be due to what Hutchins calls "overattribution": the problem arise only if the computations are attributed to the isolated mind. By reconsidering the role of the environment in cognition, then information search may re-appear tractable. In fact, sociologists of scientific knowledge have studies the further missing determinations of belief formation, which are part of what is missing to Fodor for understanding belief formation. The relevance of social studies of science for understanding the cognitive processes of scientific belief formation is made manifest when one specifies, as I tried to do, the cognitive foundations and implications of the phenomena described by sociologists of scientific knowledge. Empirical studies of scientific belief formation show that the information actually taken in consideration for problem solving is determined by the structure of the environment, especially by the accessibility or salience of beliefs and ideas. Moreover, an important part of the cognitive processes of scientific belief formation happen outside of the mind and depend on the social organisation. Thus, one partial response to the tractability problem is to consider the situated aspects of cognition, and understanding these situated aspects requires analysing the structure of the environment. The socially situated and distributed aspects of human cognition, which is shown to be important in scientific cognition, are largely due to human meta-representational capacities.

Specifying the embodiment of scientific cognition — in brains as biological organs, and in social institutions, as distributed cognitive systems — is, I hope, contributing to the projects of cognitive science and naturalised epistemology, which have the common goal of understanding the natural implementation of knowledge production. This specification also brings new insights about the relation between cognition and the evolution of scientific knowledge. The two are intimately related because scientific knowledge is a cultural phenomenon and because scientific cognition is very sensitive to culture. As human cognition is situated, it importantly changes when the environment change. The environment changes when, among other things, it is populated with new public representations and new artefacts. The evolution of science and technology has an impact on scientific cognition, which then determines further evolution. This impact is not necessarily resulting in change in the architecture of the mind; it is resulting on the framing of scientific thoughts, as determined by interpretive traditions, in the use of cognitive tools, and in the framing of the institutions of science.

Bibliography

M. Alac and E. Hutchins. I see what you are saying: Action as cognition in fmri brain mapping practice. *Journal of Cognition and Culture*, 4(3–4): 629–662, 2004. 39

S. Atran. *Cognitive Foundations of Natural History: Towards an Anthropology of Science*. Cambridge University Press, 1990. 38, 122, 162, 163, 166, 173, 405, 408

S. Atran. Folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and Brain Sciences*, 21:547–569, 1998. 109, 408

J. H. Barkow, L. Cosmides, and J. Tooby, editors. *The Adapted Mind : Evolutionary Psychology and the Generation of Culture*. Oxford University Press, 1992. ISBN 0195101073. 38, 108

B. Barnes. tural rationality: A neglected concept in the social sciences. *Philosophy of the Social Sciences*, 6:115–126, 1976. 54

B. Barnes. Social life as bootstrapped induction. *Sociology*, 17:524–45, 1983. 52, 73, 285, 294

B. Barnes and D. Bloor. Relativism, rationalism and the sociology of knowledge. In M. Hollis and S. Lukes, editors, *Raionality and Relativism*, pages 21–47. Basil Blackwell, Oxford, 1982. 95

B. Barnes, D. Bloor, and J. Henry. *Scientific Knowledge: A Sociological Analysis*. University of Chicago Press, 1996. 54, 59, 60, 68, 69, 96, 99, 165, 365, 404

M. Beller. *Quantum Dialogue*. University of Chicago Press, 1999. 25, 143

J. P. V. Bendegem and B. V. Kerkhove. The unreasonable richness of mathematics. *Journal of Cognition and Culture*, 4(3–4):525–550, 2004. 39

B. Berlin and P. Kay. *Basic color terms*. University of California Press, Berkeley, CA, 1969. 38, 90

W. Bijker, T. Hughes, and T. Pinch, editors. *The Social Construction of Technological Systems*. MIT Press, Cambridge, MA., 1987. 245

M. Blay. Deux moments de la critique du calcul infinitésimal: Michel Rolle et George Berkeley. *Revue d'Histoire des Sciences*, 39(3):223–253, 1986. 223, 234, 236

M. Bloch. Language, anthropology and cognitive science. *Man*, 26:183–198, 1991. 281

N. Block. Introduction: What is functionalism? In *Readings in philosophy of psychology*. Harvard University Press, Cambridge, MA., 1980. 291

N. Block. Qualia. In S. Guttenplan, editor, *A Companion to Philosophy of Mind*. Blackwell, Oxford, 1994. 291

D. Bloor. Institution and rule-scepticism: A reply to martin kush. *Social Studies of Science*, 34(4):593–601, August 2004a. 64

D. Bloor. Sociology of scientific knowledge. In I. Niiniluoto, M. Sintonen, and J. Wolenski, editors, *Handbook of Epistemology*, pages 919–962. Kluwer, Dordrecht, 2004b. 51

D. Bloor. *Knowledge and Social Imagery*. University Of Chicago Press, Chicago, IL, second edition: 1991 edition, September 1976. 30, 46, 49, 54, 57, 131, 198, 357, 358

D. Bloor. Polyhedra and the abominations of leviticus. *British Journal for the History of Science*, 11:245–72, 1978. 197, 230

D. Bloor. Durkheim and Mauss revisited: classification and the sociology of knowledge. *Studies in History and Philosophy of Science*, 13(4):267–297, 1982. 175, 176

D. Bloor. Ordinary human inferences as material for the sociology of knowledge. *Social Studies of Science*, 22(1):129–139, February 1992. 54

D. Bloor. What can the sociologist of knowledge say about 2+2=4? In *Mathematics, Education and Philosphy*. Falmer, London, 1994. 197

D. Bloor. Idealism and the sociology of knowledge. *Social Studies of Science*, 26(4):839–856, November 1996. 225

D. Bloor. *Wittgenstein on Rules and Institutions*. Routledge, London, 1997a. 53, 73, 285, 294, 304

D. Bloor. Remember the strong programme? *Science, Technology, & Human Values*, 22(3), Summer 1997b. 54, 56

R. Boudon. *The logic of Social Action*. Routledge and Kegan Paul, 1979. 225

P. Bourdieu. *Science de la science et réflexivité*. Raisons d'Agir, 2001. 35

P. Bourdieu. La spécificité du champ scientifique et les conditions sociales du progrès de la raison. *Sociologie et Société*, 7(1):91–118, 1975. 35

P. Bourdieu. Le champ scientifique. *Actes de la recherches en sciences sociales*, 2–3:88–104, 1976. 35

P. Bourdieu. *Outline of a Theory of Practice*. Cambridge University Press, 1977. 35

P. Bourdieu. *Homo Academicus*. Les editions de minuit, Paris, 1984. 35

R. Boyd and P. J. Richerson. *Culture and the evolutionary process*. University of Chicago Press, 1985. 225

C. B. Boyer. *The history of the calculus and its conceptual development*. Dover, New York, 1959. 223

P. Boyer. *Religion explained: the evolutionary origins of religious thought*. Basic Books, 2001. 340

M. Bradie. Assessing evolutionary epistemology. *Biology and Philosophy*, 1: 401–459, 1986. 377

E. M. Brannon and J. D. Roitman. Nonverbal representations of time and number in animals and human infants. In W. H. Meck, editor, *Functional and neural mechanisms of interval timing*, pages 143–182. CRC Press, 2003. 203

J. Bricmont and A. Sokal. Science and sociology of science: Beyond war and peace. In J. Labinger and H. Collins, editors, *The One Culture?: A Conversation about Science*, pages 27–47. University of Chicago Press, 2001. 17, 50, 62

T. L. Brown. *Making Truth: The Roles of Metaphor in Science*. University of Illinois Press, 2003. 166

D. T. Campbell. Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review*, 67:380–400, 1960. 376, 379

D. T. Campbell. Evolutionary epistemology. In P. A. Schlipp, editor, *The philosophy of Karl Popper*, pages 413–63. Open Court, 1974. 312, 376, 380

S. Carey. Cognitive foundations of arithmetic: Evolution and ontogenesis. *Mind and Language*, 16:37–55, 2001. 206

S. Carey. *Conceptual Change in Childhood. Cambridge, MA: Bradford Books, MIT Press.* Bradford Books, MIT Press, 1985. 400

S. Carey. On the origin of causal understanding. In D. P. D. Sperber and A. J. Premack, editors, *Causal cognition: a multidisciplinary debate*, pages 268–308. Oxford University Press, 1995. 407

S. Carey and S. Johnson. Metarepresentation. In D. Sperber, editor, *Metarepresentation and conceptual change: Evidence from Williams Syndrome*, pages 225–264. Cambridge University Press, 2000. 400

S. Careyand and E. Spelke. Domain-specific knowledge and conceptual change. In L. Hirschfeld and S. Gelman, editors, *Mapping the Mind: Domain Specificity in Cognition and Culture*. Cambridge University Press, Cambridge, 1994. 400, 406, 407

P. Carruthers. The roots of scientific reasoning: infancy, modularity and the art of tracking. In P. Carruthers, M. Siegal, and S. P. Stich, editors, *The Cognitive Basis of Science*, chapter 4, pages 73–97. Cambridge University Press, 2002. 118, 119

P. Carruthers. Moderately massive modularity. In A. O'Hear, editor, *Minds and persons*, pages 67–90. Cambridge University Press, 2003. 399

P. Carruthers and A. Chamberlain, editors. *Evolution and the human mind: Modularity, language and meta-cognition*. Cambridge University Press, 2000. 108

P. Carruthers, S. Stich, and M. Siegal. Introduction: What makes science possible? In P. Carruthers, S. Stich, and M. Siegal, editors, *The Cognitive Basis of Science*, chapter 1. Cambridge University Press, 2002. 4, 39

P. Carruthers, S. Laurence, and S. Stich, editors. *The Innate Mind: Structure and Contents*. Oxford University Press, USA, July 2005. 108

J. Cavailles. *Remarques sur la formation de la théorie abstraite des ensembles*. Hermann et Cie, 1938. 232

P. Churchland. Perceptual plasticity and theoretical neutrality: A reply to jerry fodor. *Philosophy of Science*, 55(2):167–187, 1988. 38

A. Clark. *Being There: Putting Mind, Body and Brain Together Again*. MIT Press, 1997. 276, 277, 289, 295, 303

F. Clement, M. Koenig, and P. Harris. The ontogenesis of trust. *Mind and Language*, 19(4):360–379, September 2004. doi: 10.1111/j.0268-1064.2004.00263.x. 324

C. A. J. Coady. *Testimony : A Philosophical Study*. Oxford University Press, USA, March 1995. ISBN 0198235518. 319, 323

H. M. Collins. *Changing Order: Replication and Induction in Scientific Practice*. Sage Publication, London, 1992. 59

D. L. Craig, N. J. Nersessian, and R. Catrambone. Perceptual simulation in analogical problem solving. In L. Magnani and N. J. Nersessian, editors, *Model-Based Reasoning: Science, Technology, and Values*, pages 167–191. Kluwer Academic-Plenum Publishers, 2002. 270

D. D. Cummins and C. Allen. Introduction. In D. D. Cummins and C. Allen, editors, *The Evolution of Mind*. Oxford University Press, 1998. 395

L. Daston, editor. *Biographies of scientific objects*. University of Chicago Press, 2000. 415

R. Dawkins. *The Selfish Gene*. Oxford University Press, Oxford, 1976. 151, 312

H. De Cruz. How do cultural numerical concepts build upon an evolved number sense? In L. B. B.G. Bara and M. Bucciarelli, editors, *Proceedings of the XXVII Annual Conference of the Cognitive Science Society*, pages 565–570. Lawrence Erlbaum, 2005. 238

S. Dehaene. How a primate brain comes to know some mathematical truths. Rencontre internationale de la fondation IPSEN "Neurobiologie des valeurs humaines, Paris, January 24, 2005., 2005. 190

S. Dehaene. *The Number Sense : How the Mind Creates Mathematics*. Oxford University Press, December 1999. ISBN 0195132408. 187, 202, 203, 204

S. Dehaene, G. Dehaene-Lambertz, and L. Cohen. Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, 21: 355–361, 1998. 203

D. C. Dennett. *Brainstorms*, chapter Intentional systems. MIT Press, a Bradford Book, 1978. 134

M. Donald. *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Harvard University Press, 1991. 303

S. M. Downes. Agents and norms in the new economics of science. *Philosophy of the Social Sciences*, 31(2):224 – 238, 2001. 17, 20

K. Dunbar. How scientists really reason: Scientific reasoning in real-world laboratories. In R. S. . J. Davidson, editor, *Mechanisms of insight*, pages 365–395. MIT press, Cambridge MA, 1995. 15

K. Dunbar and I. Blanchette. The invivo/invitro approach to cognition: the case of analogy. *Trends in Cognitive Sciences*, 5:334–339, 2001. 15

K. Dunbar and J. A. Fugelsang. Causal thinking in science: How scientists and students interpret the unexpected. In M. Gorman, R. Tweney, D. Gooding, and A. Kincannon, editors, *New Directions in Studies of Scientific and Technological Thinking*. Lawrence Elbraum, Mahwah, NJ, 2005. 15

E. L. Eisenstein. *The Printing Press as an Agent of Change (Volumes 1 and 2 in One)*. Cambridge University Press, September 1980. ISBN 0521299551. 252

R. Ellen. From ethno-science to science, or 'what the indigenous knowledge debate tells us about how scientists define their project'. *Journal of Cognition and Culture*, 4(3–4):409–550, 2004. 37

J. L. Elman, E. A. Bates, M. H. Johnson, and K. P. Annette Karmiloff-Smith, Domenico Parisi. *Rethinking Innateness: A Connectionist Perspective on Development*. MIT Press, 1996. 112

P. Engel. *La norme du vrai, philosophie de la logique*. Gallimard, Paris, 1989. 191

A. Erana and S. F. Martinez. The heuristic structure of scientific knowledge. *Journal of Cognition and Culture*, 4(3–4):701–730, 2004. 39, 411

J. S. B. T. Evans. In two minds: dual-process account of reasoning. *Trends in Cognitive Sciences*, 7(10):454–459, 2003. 139, 170

G. J. Feist and M. E. Gorman. The psychology of science: review and integration of a nascent discipline. *Review of General Psychology*, 2(1): 3–47, 1998. 29

J. Fodor. *The Mind Doesn't Work That Way*. MIT Press, Cambridge, Mass., 2000. 394, 397, 407

J. Fodor. *The Modularity of Mind*. A Bradford Book. MIT Press, 1983. 397

J. Fodor. *In critical conditions: polemical essays on cognitive science and the philosophy of mind*. MIT Press, A Bradford Book, 1998. 133

B. Fontenelle. *Éléments de la géométrie de l'infini*. Klincksieck [Edition 1995], 1727. 227

E. G. Freedman. Understanding scientific discourse: A strong programme for the cognitive psychology of science. *Theory and Review in Psychology*, Online review: http://gemstate.net/susan/Eric.htm:accessed on 15/8/06, 1997. 4

G. Frege. *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Verlag Hermann Pohle, Breslau, 1884. 181

G. Frege. *Grundgesetze der Arithmetik*. Verlag Hermann Pohle, Jena, 1893. 181

F. Fritsch. *The four-color theorem: history, topological foundations, and idea of proof*. Springer, 1998. 329

S. Fuller, T. Shinn, M. de Mey, and S. Woolgar, editors. *The Cognitive Turn: Sociological and Psychological Perspectives on Science*. Springer, 1989. 17

P. Galison. *How Experiments End*. University of Chicago Press, 1987. 88

C. R. Gallistel and R. Gelman. Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, 4:59–65, 2000. 187, 208

C. R. Gallistel and R. Gelman. Mathematical cognition. In K. Holyoak and R. Morrison, editors, *The Cambridge handbook of thinking and reasoning*, pages 559–588. Cambridge University Press, 2005. 203, 206, 207, 208

C. R. Gallistel, R. Gelman, and S. Cordes. The cultural and evolutionary history of the real numbers. In S. Levinson and P. Jaisson, editors, *Culture and evolution*. MIT Press, 2005. 181, 197, 215

M. Gibbons, C. Limoges, H. Nowotny, S. Schwartzman, P. Scott, and M. Trow. *The New Production of Knowledge: The dynamics of science and research in contemporary societies*. Sage Publications Ltd, 1994. 353

R. Giere. Scientific cognition as distributed cognition. In P. Carruthers, S. Stich, and M. Siegal, editors, *The Cognitive Basis of Science*, chapter 15, pages 285–299. Cambridge University Press, Cambridge, UK, 2002a. 251, 258, 300

R. Giere. Computation and agency in scientific cognition. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, Boston, MA., August 2003. 251, 295, 296

R. Giere. The cognitive structure of scientific theories. *Philosophy of Science*, 61(2):276–296, 1994. 39, 53

R. Giere and B. Moffatt. Distributed cognition: Where the cognitive and the social merge. *Social Studies of Science*, 33:301–310, April 2003. 40, 251, 257, 258, 286, 300, 325

R. N. Giere. Distributed cognition in epistemic cultures. *Philosophy of Science*, pages 637–644, December 2002b. 258, 351

R. N. Giere. Models as parts of distributed cognitive systems. In L. Magnani and N. Nersessian, editors, *Model Based Reasoning: Science, Technology, Values*, pages 227–41. Kluwer, 2002c. 270

R. N. Giere. The problem of agency in scientific distributed cognitive systems. *Journal of Cognition and Culture*, 4(3–4):759–74, 2004. 259, 268

R. N. Giere. *Explaining Science: A Cognitive Approach*. Science and Its Conceptual Foundations. University Of Chicago Press, 1988. 23, 26, 39, 169, 257

R. N. Giere. The cognitive construction of scientific knowledge (response to pickering). *Social Studies of Science*, 22(1):95–107, 1992. 26, 72, 392

G. Gigerenzer and R. Selten, editors. *Bounded rationality: the adaptive toolbox*. MIT Press, 2001. 137

G. Gigerenzer, P. M. Todd, and the ABC Research Group. *Simple Heuristics That Make Us Smart*. Oxford University Press, New York, 1999. 129, 137, 381, 382

A. I. Goldman. *Epistemology and Cognition*. Harvard University Press, October 1986. ISBN 0674258967. 19, 86, 131

A. I. Goldman. *Liaisons: Philosophy meets the cognitive and the social sciences*. MIT Press, 1992. 19

A. I. Goldman. Psychological, social, and epistemic factors in the theory of science. *PSA: Proceedings of the Biennal Meeting of the Philosophy of Science Association*, 2:277–286, 1994. 19, 24

A. I. Goldman. *Knowledge in a Social World*. Oxford University Press, February 1999. ISBN 0198238207. 19, 22, 319, 323, 324

D. Gooding. Seeing the forest for the trees: Visualization, cognition and scientific inference. In M. Gorman, D. Gooding, R. Tweney, and A. Kincannon, editors, *Scientific and Technological Thinking*. Lawrence Erlbaum Publishers, 2005. 34

D. Gooding. *Experiment and the making of meaning: Human agency in scientific observation and experiment*. Kluwer, 1990. 31, 337

D. C. Gooding. Cognition, construction and culture: Visual theories in the sciences. *Journal of Cognition and Culture*, 4(3–4):551–594, 2004. 34

J. Goody. *The Domestication of the Savage Mind*. Cambridge University Press, November 1977. ISBN 0521292425. 231, 252, 303, 346

A. Gopnik. The scientist as child. *Philosophy of Science*, 63(4):485–514, December 1996. 139, 405

A. M. Gopnik and A. N. Meltzoff. *Words, Thoughts, and Theories*. MIT Press, 1997. 118

M. Gorman. Heuristics in technoscientific thinking. *Behavioral and Brain Sciences*, 23:752, 2000. 403

M. Gorman. *Simulating science: Heuristics, mental models and technoscientific thinking*. Indiana University Press, 1992. 31

N. Guicciardini. Newton's method and Leibniz's calculus. In H. N. Jahnke, editor, *A History of Analysis*. American Mathematical Society, 2003. 211, 234

N. Guicciardini. Tree traditions in the calculus: Newton, leibniz and lagrange. In I. Grattan-Guinness, editor, *Companion encyclopedia of the history and philosophy of the mathematical sciences*, volume 1, pages 308–17. Routledge, London, 1994. 210

D. Guin and L. Trouche. The complex process of converting tools into mathematical instruments: The case of calculators. *International Journal of Computers for Mathematical Learning*, 3(3):195–227, October 1998. 347

I. Hacking. *The social construction of what?* Harvard University Press, 1999. 124, 391

J. Hardwig. Epistemic dependence. *Journal of Philosophy*, 82(7):335–349, 1985. 101, 319

J. Hardwig. The role of trust in knowledge. *Journal of Philosophy*, 88(12): 693–708, 1991. 101, 319, 347, 362

L. A. Hirschfeld and S. A. Gelman, editors. *Mapping the Mind : Domain Specificity in Cognition and Culture*. Cambridge University Press, April 1994. ISBN 0521429935. 38, 108

J. Hollan, E. Hutchins, and D. Kirsh. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Trans. Comput.-Hum. Interact.*, 7(2):174–196, June 2000. ISSN 1073-0516. doi: 10.1145/353485.353487. 337

M. Hollis and S. Lukes, editors. *Rationality and Relativism*. Basil Blackwell, Oxford, 1982. 95

R. Horton. African traditional thought and western science. In B. Wilson, editor, *Rationality*. Blackwell, Oxford, 1970. 93

D. L. Hull. *Science and Selection: Essays on Biological Evolution and the Philosophy of Science*. Cambridge University Press, 2001. 134, 386, 388

D. L. Hull. *Science as a process: an evolutionary account of the social and conceptual development of science*. The University of Chicago Press, 1988. 66, 386

E. Husserl. *Logische Untersuchungen. Erste Teil: Prolegomena zur reinen Logik*. Kluwer Academic Publishers, 1900. 181

E. Hutchins. *Cognition in the Wild*. The MIT Press, (Bradford Books), 1995. 9, 40, 118, 247, 253, 256, 278, 279, 280, 293, 302, 303, 321, 344, 348, 353, 398

H. N. Jahnke, editor. *A history of analysis*. American Mathematical Society, 2003. 223

P. Johnson-Laird. *Mental Models*. Harvard University Press, Cambridge, MA, 1983. 39

P. Johnson-Laird and R. Byrne. *Deduction*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1991. 187

D. Kahneman, P. Slovic, and A. Tversky, editors. *Judgment Under Uncertainty*. Cambridge University Press, 1982. 20, 170

A. Karmiloff-Smith. *Beyond Modularity: A developmental perspective on cognitive science*. MIT Press, 1992. 396

K.-M. Kim. Review: Natural versus normative rationality: reassessing the strong programme in the sociology of knowledge. *Social Studies of Science*, 24(2):391–403, 1994. 131

P. Kitcher. Reviving the sociology of science. *Philosophy of Science*, 67:33–44, September 2000. 18, 20, 72

P. Kitcher. *The Nature of Mathematical Knowledge*. Oxford University Press, Oxford and New York, 1984. 215, 337

P. Kitcher. The naturalists return. *Philosophical Review*, 101:53–114, 1992. 19, 22

P. Kitcher. *The Advancement of Science*. Oxford University Press, New York, 1993. 19, 23

P. Kitcher. A plea for science studies. In N. Koertge, editor, *A House Build on Sand: Exposing Postmodernist Myths About Science*. Oxford University Press, New York, 1998. 72

D. Klahr and H. Simon. Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin*, 125(5):524–543, 1999. 15

K. Knorr Cetina. *Epistemic Cultures: How the Sciences Make Knowledge*. Harvard University Press, Cambridge, MA, 1999. 258, 351, 358

M. A. Koenig and P. L. Harris. The role of social cognition in early trust. *Trends in Cognitive Sciences*, 9(10):457–459, October 2005. doi: 10.1016/j.tics.2005.08.006. 324

M. E. Kronfeldner. Darwinian 'blind' hypothesis formation revisited. *British Journal for the Philosophy of Science*, to appear. 379

D. Kuhn. Is good thinking scientific thinking? In D. R.Olson and N. Torrance, editors, *Modes of thought: explorations in culture and cognition*. Cambridge University Press, 1996. 403

E. M. Kurz and R. D. Tweney. The practice of mathematics and science : From calculus to the clothesline problem. In M. Oaksford and N. Chater, editors, *Rational models of cognition*, pages 415–438. Oxford University Press, 1998. 214

E. Kurz-Milcke, N. J. Nersessian, and W. C. Newstetter. What has history to do with cognition? interactive methods for studying research laboratories. *Journal of Cognition and Culture*, 4(3-4):663–700, 2004. 39, 251, 263, 270, 272, 286, 301, 315, 323

M. Kusch. Rule-scepticism and the sociology of scientific knowledge: The bloor-lynch debate revisited. *Social Studies of Science*, 34(4):571–91, August 2004. 64

I. Lakatos. *Proofs and Refutations : The Logic of Mathematical Discovery*. Cambridge University Press, January 1976. ISBN 0521290384. 167, 229, 335

I. Lakatos. *Mathematics, science and epistemology — Philosophical papers*, volume 2, chapter Cauchy and the continuum: the significance of non-standard analysis for the history and philosophy of mathematics, pages 43–60. Cambridge University Press, 1978. 197, 212, 213, 220

G. Lakoff and R. Nuñez. *Where mathematics comes from: how the embodied mind brings mathematics into being*. Basic Book, 2000. 208, 214, 215, 217, 337

P. Langley, G. L. Bradshaw, and H. A. Simon. Bacon.5: The discovery of conservation laws. In *IJCAI*, pages 121–126, 1981. 28

B. Latour. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press, 2005. 325

B. Latour. Visualisation and cognition: Thinking with eyes and hands. *Knowledge and Society*, 6:1–40, 1986. 15, 117, 252, 254, 255, 258, 260

B. Latour. *Science in Action: How to Follow Scientists and Engineers Through Society*. Harvard University Press, October 1987. ISBN 0674792912. 28, 61, 62, 63, 119

B. Latour. *The Pasteurization of France*. Harvard University Press, Cambridge, MA, 1993. 62, 325

B. Latour. Cogito ergo sumus: Review of ed hutchins' 'cognition in the wild'. *Mind, Culture and Activity*, 3(1):54–63, 1996. 251, 252

B. Latour. *Pandora's Hope: Essays on the Reality of Science Studies*. Harvard University Press, Cambridge, MA, 1999a. 127, 129, 258, 260

B. Latour. For david bloor ... and beyond : A reply to david bloor's "anti-latour". *Studies in the history and philosophy of science*, 30(1):113–129, March 1999b. 75

B. Latour and S. Woolgar. *Laboratory Life: The construction of Scientific Facts*. Princeton University Press, 1986. 28, 63, 119

L. Laudan. Relativism, naturalism and reticulation. *Synthese*, 71:221–34, 1987. 17

L. Laudan. *Science and Relativism*. University of Chicago Press, 1990. 4

J. Lave. *Cognition in Practice: mind, mathematics and culture in everyday life*. Cambridge University Press, 1988. 33, 40, 201

A. Lillard. Ethnopsychologies: Cultural variations in theories of mind,. *Psychological Bulletin*, 123(1):3–32, January 1998. 91

J. Lipton and E. Spelke. Origins of number sense: large-number discrimination in human infants. *Psychological Science*, 14:396–401, 2003. 204

E. Livingston. *The Ethnomethodological Foundations of Mathematics*. Routledge and Kegan Paul Books Ltd, 1986. 201

O. Lizardo. The cognitive origins of bourdieu's habitus. *Journal for the Theory of Social Behaviour*, 34(4), 2004. 35

H. E. Longino. *The Fate of Knowledge*. Princeton University Press, 2001. 26, 27

S. Lukes. Relativism in its place. In M. Hollis and S. Lukes, editors, *Rationality and Relativism*. Basil Blackwell, 1982. 93

M. Lynch. Ethnomethodology and the logic of practice. In K. Knorr-Cetina, E. von Savigny, and T. Schatzki, editors, *The Practice Turn in Contemporary Theory*. Routledge, London, 2001. 64

M. Lynch. Extending wittgenstein: the pivotal move from epistemology to the sociology of science. In A. Pickering, editor, *Science as Practice and Culture Chicago*, pages 215–65. University of Chicago Press, 1992. 63, 64

D. MacKenzie. Negotiating arithmetic, constructing proof: The sociology of mathematics and information technology. *Social Studies of Science, Vol. 23*, 23(1):37–65, February 1993. 363

D. MacKenzie. *Knowing Machines: Essays on Technical change*. MIT Press, Cambridge, MA., 1996. 305, 306

D. MacKenzie. The sociology of a mathematical proof. *Social Studies of Science*, 29(1):7–70, February 1999. 328, 334, 358, 359, 362

J. Macnamara. *A Border Dispute, the Place of Logic in Psychology*. MIT Press, Cambridge (MA), 1986. 183, 184

J. Macnamara and G. E. Reyes, editors. *The Logical Foundations of Cognition*. Oxford University Press, 1994. 184

P. Maddy. Perception and mathematical intuition. *The Philosophical Review*, 89(2):163–196, 1980. 188, 190

P. Maddy. The roots of contemporary platonism. *The Journal of Symbolic Logic*, 54(4):1121–1144, 1989. 190

P. Maddy. *Realism in Mathematics*. OUP, Oxford, 1990. 188

P. Maddy. Set theoretic naturalism. *The Journal of Symbolic Logic*, 61(2): 490–514, 1996. 188

L. Magnani and N. Nersessian, editors. *Model-Based Reasoning: Science, Technology, Values*. Kluwer Academic/Plenum Publishers, New York, 2002. 39

R. Mallon and S. Stich. The odd couple. *Philosophy of Science*, 67:133–154, 2000. 122

P. Mancosu. The metaphysics of the calculus: A foundational debate in the paris academy of sciences, 1700-1706. *Historia Mathematica*, 16:224–248, 1989. 223, 228, 234, 235

J. G. March. *Decisions and Organizations*. Basil Blackwell Ltd, Oxford, 1988. 279

G. Marcus. *The Birth of the Mind: How a Tiny Number of Genes Creates the Complexities of Human Thought*. Basic Books, New York, 2004. 120

J. McClelland, D. E. Rumelhart, and the PDP Research Group. *Parrallel Distributed Processing: Exploration in the Microstructure of Cognition*. MIT Press, 1986. 33, 343

H. Mehrtens. Irresponsible purity: The political and moral structure of mathematical sciences in the national socialist state. In M. Renneberg and M. Walker, editors, *Science, Technology and National Socialism*. Cambridge University Press, 1994. 227

H. Mercier. Some ideas to study the evolution of mathematics. In N. Gontier, J. P. van Bendegem, and D. Aerts, editors, *Evolutionary epistemology, language and culture: a nonadaptationist systems theoretical approach*. Springer, 2006. 190

S. Mithen. Human evolution and the cognitive basis of science. In P. Carruthers, S. Stich, and M. Siegal, editors, *The cognitive basis of science*. Cambridge University Press, 2002. 139

J. Mundale and W. Bechtel. Integrating neuroscience, psychology, and evolutionary biology through a teleological conception of function. *Minds and Machines*, 6(4):481–505, November 1996. 291

R. Nelson and S. Winter. *An Evolutionary Theory of Economic Change*. Harvard University Press, Boston, MA, 1982. 294

N. Nersessian. In the theoretician's laboratory: Thought experimenting as mental modeling. *PSA*, 2:291–301, 1992a. 39, 400

N. J. Nersessian. The cognitive basis of model-based reasoning in science. In P. Carruthers, S. Stich, and M. Siegal, editors, *The Cognitive Basis of Science*, pages 133–153. Cambridge University Press, 2002a. 257, 402

N. J. Nersessian. Kuhn, conceptual change, and cognitive science. In T. Nichols, editor, *Thomas Kuhn*, Contemporary Philosophers in Focus Series, pages 178–211. Cambridge University Press, 2002b. 257, 392

N. J. Nersessian. Maxwell and "the method of physical analogy": Model-based reasoning, generic abstraction, and conceptual change. In D. Malament, editor, *Essays in the History and Philosophy of Science and Mathematics*, pages 129–166. Open Court, Lasalle, Il, 2002c. 31

N. J. Nersessian. Interpreting scientific and engineering practices: Integrating the cognitive, social, and cultural dimensions. In M. Gorman, R. Tweney, D. Gooding, and A. Kincannon, editors, *Scientific and Technological Thinking*, pages 17–56. Lawrence Erlbaum Publishers, 2005. 27, 251, 262, 265

N. J. Nersessian. The cognitive-cultural systems of the research laboratory. *Organization Studies*, 27(1):125–145, 2006. 31, 33, 257, 272

N. J. Nersessian. *Faraday to Einstein: Constructing Meaning in Scientific Theories*. Kluwer, Dordrecht, 1984. 31, 257, 337

N. J. Nersessian. How do scientists think? capturing the dynamics of conceptual change in science. In R. N. Giere, editor, *Cognitive Models of Science*, pages 3–45. University of Minnesota Press, Minneapolis, MN, 1992b. 31, 257

N. J. Nersessian. Opening the black box: Cognitive science and history of science. *Osiris, 2nd Series*, 10:194–211, 1995. 16, 31, 32, 257, 263, 336, 392

N. J. Nersessian. Model-based reasoning in conceptual change. In L. Magnani, N. J. Nersessian, and P. Thagard, editors, *Model-Based Reasoning in Scientific Discovery*, pages 5–22. Kluwer Academic/Plenum Publishers, New York, 1999. 166, 257

A. Newell, J. C. Shaw, and H. Simon. Elements of a theory of human problem solving. *Psychological Review*, 65:151–66, 1958. 380

W. Newton-Smith. *The Rationality of Science*. Routledge and Kegan Paul, 1981. 131

R. Nisbett and L. Ross. *Human Inferences: Strategies and Shortocomeings of Social Judgment*. Prentice-Hall, Engewood Cliffs, N.J., 1980. 90

M.-P. Noel. Numerical cognition. In B. Rapp, editor, *The handbook of cognitive neuropsychology: What deficits reveal about the human mind*, pages 495–518. Psychology Press, 2001. 204

F. J. Odling-Smee, K. N. Laland, and M. W. Feldman. *Niche Construction: the Neglected Process in Evolution*. Priceton University Press, Princeton, New Jersey, 2003. 312

D. R. Olson. *The world on paper: the conceptual and cognitive implications of writing and reading*. Cambridge University Press, 1994. 303

G. Origgi. Is trust an epistemological notion ? *Episteme*, 1(1):61–72, 2004. 319

G. Origgi, D. Sperber, and C. Heintz, editors. *Rethinking Interdisciplinarity*, 2004. CNRS, Insitut Jean Nicod. URL `http://www.interdisciplines.org/interdisciplinarity`. 352

J. Piaget. *La Construction du réel chez l'enfant*. Delachaux et Niestlé, Neuchâtel, 1937. 392

P. Pica, C. Lemer, V. Izard, and S. Dehaene. Exact and approximate arithmetic in an Amazonian indigene group. *Science*, 306:499–503, 2004. 204

A. Pickering. Philosophy naturalized a bit. *Social Studies of Science*, 21(3): 575–584, aug 1991. ISSN 0306-3127. 26, 257

A. Pickering, editor. *Science as Practice and Culture*. University of Chicago Press, 1992. 59, 65, 245

S. Pinker. *The Blank Slate: The Modern Denial of Human Nature*. Viking Adult, 2002. 116

A. Plantinga. *Warrant and Proper Function*. Oxford University Press, 1993. 133

H. Poincaré. *La science et l'hypothèse*. Flammarion [édition de 1968], 1902. 231, 232

D. A. Poling and E. M. Evans. Religious belief, scientific expertise, and folk ecology. *Journal of Cognition and Culture*, 4(3–4):485–524, 2004. 39

K. Popper. Replies to my critics. In P. Schilpp, editor, *The philosophy of Karl Popper*. Open Court Press, 1974. 383

W. Quine. *Ontological Relativity and Other Essays*, chapter Epistemology naturalized. Columbia University Press, 1969. 11

N. Quinn. Convergent evidence for a cultural model of american marriage. In D. Holland and N. Quinn, editors, *Cultural Models in Language and Thought*. Cambridge University Press, Cambridge, 1987. 39

M. Resnik. *Turtles, Termites, and Traffic Jams: Exploration in Masssively Parallel Microworlds*. MIT Press, 1994. 278, 288

R. M. Roberts. *Serendipity: Accidental Discoveries in Science*. John Wiley & Sons Inc, 1989. 384

A. Robinet. Le groupe malebranchiste introducteur du calcul infinitésimal en france. *Revue d'Histoire des Sciences*, 13:287–308, 1960. 223

W.-M. Roth. Emergence of graphing practices in scientific research. *Journal of Cognition and Culture*, 4(3–4):595–628, 2004. 34, 39

W.-M. Roth and G. Bowen. When are graphs ten thousand words worth? an expert/expert study. *Cognition and Instruction*, 21:429–473, 2003. 34

R. Samuels. Nativism in cognitive science. *Mind and Language*, 17:233–265, 2002. 112

D. Schmandt-Besserat. *Before Writing*, volume Volume 1. University of Texas Press, 1992. 346

Science et Vie. No 1013 - l'intelligence dévoile enfin sa vrai nature : toute pensée est un calcul!, February 2002. 187

S. Shapin. *A Social History of Truth : Civility and Science in Seventeenth-Century England*. Science and Its Conceptual Foundations. University Of Chicago Press, November 1995. ISBN 0226750191. 319

S. Shapin and S. Schaffer. *Levianthan and the Air-Pump: Hobbes, Boyle and the experimental life*. Princeton University Press, 1985. 175

H. Simon. The architecture of complexity. In *The Sciences of the Artificial*, pages 1983–216. MIT Press, Cambridge, Massachusetts, 1996. 395

D. K. Simonton. Creativity as blind variation and selective retention: Is the creative process darwinian. *Psychological Inquiry*, 10:309–28, 1999. 382

P. Slezak. Scientific discovery by computers as empirical refutation of the strong programme. *Social Studies of Science*, 9(4):563–600, November 1989. 4, 17, 57

H. Small and B. C. Griffith. The structure of scientific literatures i: Identifying and graphing specialties. *Science Studies*, 4(1):17–40, January 1974. 351

E. Sober. Innate knowledge. In *The Routledge Encyclopedia of Philosophy*, volume 4, pages 794–797. Routledge, 1999. 112

M. Solomon. *Social Empiricism*. MIT Press, 2006. 26

E. S. Spelke. Nativism, empiricism, and the origins of knowledge. *Infant Behavior and Development*, 21(2):181–200, 1998. 108

D. Sperber. Metarepresentations in an evolutionary perspective. In D. Sperber, editor, *Metarepresentations: a multidisciplinary perspective*, pages 117–137. Oxford University Press, 2000. 141

D. Sperber. Conceptual tools for a natural science of society and culture. *Proceedings of the British Academy*, 111:297–317, 2001a. 148, 149, 158

D. Sperber. Conceptual tools for a natural science of society and culture. In *Proceedings of the British Academy*, volume 111 of *(Radcliffe-Brown Lecture in Social Anthopology 1999)*, pages 297–317, 2001b. 289

D. Sperber. An evolutionary perspective on testimony and argumentation. *Philosophical Topics*, 29:401–413, 2001c. 141

D. Sperber. In defense of massive modularity. In E. Dupoux, editor, *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler*, pages 47–57. MIT Press, 2002. 399

D. Sperber. Modularity and relevance: How can a massively modular mind be flexible and context-sensitive? In P. Carruthers, S. Laurence, and S. Stich, editors, *The Innate Mind: Structure and Content*. Oford University Press, USA, 2005. 341, 399

D. Sperber. Why a deep understanding of cultural evolution is incompatible with shallow psychology. In *Roots of Human Sociality: Culture, Cognition and Interaction*. Berg Publishers, 2006a. 80, 149, 151, 275

D. Sperber. Why a deep understanding of cultural evolution is incompatible with shallow psychology. In N. Enfield and S. Levinson, editors, *Roots of Human Sociality*, pages 431–449. Berg, 2006b. 256

D. Sperber. *Rethinking Symbolism*. Cambridge University Press, 1975. 162, 166, 170

D. Sperber. Apparently irrational beliefs. In M. Hollis and S. Lukes, editors, *Rationality and Relativism*, pages 149–180. Basil Blackwell, 1982. 93, 96, 99, 102, 103

D. Sperber. Anthropology and psychology: Towards an epidemiology of representations. *Man*, 20:73–89, 1985. Reprinted in Sperber (1996a, chap. 3). 148

D. Sperber. *Explaining Culture: A Naturalistic Approach*. Blackwell Publishers, September 1996a. ISBN 0631200452. 38, 121, 150, 154, 160, 269, 281, 312, 341, 387, 408, 443

D. Sperber. Author's presentation of Writing Culture. available online at www.dan.sperber.com/exp-cult.htm (last accessed 06/01/07), 1996b. URL `www.dan.sperber.com/exp-cult.htm`. 41

D. Sperber. Inttuitive and reflective beliefs. *Mind and Language*, 12(1):67–83, 1997a. 98, 101, 408

D. Sperber. Individualisme méthodologique et cognitivisme. In R. Boudon, F. Chazel, and A. Bouvier, editors, *Cognition et sciences sociales*, pages 123–136. Presse Universitaires de France, Paris, 1997b. 104

D. Sperber and N. Claidière. Why modeling cultural evolution is still such a challenge. *Biological Theory*, 1(1):20–22, 2006. 387, 388

D. Sperber and N. Claidière. Defining and and explaining culture (comments on Richerson and Boyd, Not by genes alone). *Biology and Philosophy*, 2007. 200, 237, 238

D. Sperber and L. Hirschfeld. The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences*, 8(1):40–46, 2004. 96, 109, 121, 155, 156, 157

D. Sperber and D. Wilson. The mapping between the mental and the public lexicon. In P. Carruthers and J. Boucher, editors, *Thought and language*, pages 184–200. Cambridge University Press, 1998. 209

D. Sperber and D. Wilson. *Relevance: communication and cognition*. Harvard University Press, Cambridge, MA, USA, 1986. ISBN 0-674-75476-1. 153, 174, 341, 383, 413

D. Sperber, F. Cara, and V. Girotto. Relevance theory explains the selection task. *Cognition*, 57:31–95, 1995. 178

D. Sperber, D. Premack, and A. J. Premack, editors. *Causal Cognition : A Multidisciplinary Approach*. (A Fyssen Foundation Symposium) Oxford University Press, USA, November 1996. 108

M. Spranzi. Galileo and the mountains of the moon: Analogical reasoning, models and metaphors in scientific discovery. *Journal of Cognition and Culture*, 4(3–4):451–484, 2004. 51, 413

E. Stein and P. Lipton. Where guesses come from: Evolutionary epistemology and the anomaly of guided variation. *Biology and Philosophy*, 4: 33–56, 1989. 385

R. Sternberg. Darwinian creativity as a conventional religious faith. *Psychological Inquiry*, 10:357–59, 1999. 382

S. P. Stich. *The Fragmentation of Reason: preface to a pragmatic theory of cognitive evaluation*. MIT Press (A Bradford Book), Cambridge, MA., 1990. 85, 130, 132

P. Thagard. Societies of minds: Science as distributed computing. *Studies in History and Philosophy of Science*, 24:49–67, 1993. 251, 257

M. Tomasello. *The Cultural Origins of Human Cognition*. Harvard University Press, March 2001. ISBN 0674005821. 312

M. Tomasello, M. Carpenter, J. Call, T. Behne, and H. Moll. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28:675–691, 2005. 313

J. Tooby and L. Cosmides. The psychological foundations of culture. In J. H. Barkow, L. Cosmides, and J. Tooby, editors, *The Adapted Mind: evolutionary psychology and the generation of culture*, pages 19–136. Oxford University Press, 1992. 115

A. Tversky and D. Kahneman. Extentional versus intuitive reasoning: the conjunction fallacy in probability judgement. *Psychological Review*, 90(4): 293–315, 1983. 135

R. Tweney. Faraday's discovery of induction: A cognitive approach. In D. Gooding and F. A. J. L. James, editors, *Faraday Rediscovered: Essays on the Live and Work of Michael Faraday*, pages 189–210. Stockton Press, 1985. 31

R. Tweney. Faraday's notebooks: The active organization of creative science., 26,. *Physics Education*, 26:301–306, 1991. 31, 337

T. Tymoczko. The four-color problem and its philosophical significance. *The Journal of Philosophy*, 76(2):57–83, February 1979. 333, 362, 367

J. P. van Bendegem and B. van Kerkhove. The unreasonable richness of mathematics. *Journal of Cognition and Culture*, 4(3–4):525–549, 2004. 215, 337

E. Wigner. The unreasonable effectiveness of mathematics in the natural sciences. *Communications in Pure and Applied Mathematics*, 13(1):1–14, 1960. 191

B. R. Wilson, editor. *Rationality*. Blackwell, 1970. 37

R. Wilson. *Four colors suffice : how the map problem was solved*. Princeton university press, 2002. 329

W. Wimsatt. Heuristics refound. *Behavioral and Brain Sciences*, 23:766–7, 2000. 404

S. Woolgar. Representation, cognition and self: What hope for an integration of psychology and sociology. In S. Fuller, M. D. Mey, T. Shinn, and S. Woolgar, editors, *The Cognitive Turn: Sociological and Psychological Perspectives on Science*, pages 201–225. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1989. 61, 62

K. Wynn. Psychological foundations of number: numerical competence in human infants. *Trends in Cognitive Sciences*, 2:296–303, 1998. 204, 205

J. Zhang. External representations in complex information processing tasks. In A. Kent, editor, *Encyclopedia of Library and Information Science*, volume 68. Marcel Dekker, Inc., 2001. 346

J. Zhang. The nature of external representations in problem solving. *Cognitive Science: A Multidisciplinary Journal*, 21(2):179–217, 1997. 346