THE PONTIFICAL ACADEMY OF SCIENCES

Acta

18

PATHS OF DISCOVERY

Plenary Session
5-8 November 2004

PONTIFICIA ACADEMIA SCIENTIARVM

VATICAN CITY
2006

PATHS OF DISCOVERY

Address:

The Pontifical Academy of Sciences
Casina Pio IV, 00120 Vatican City
Tel: +39 0669883195 / Fax: +39 0669885218
Email: academy.sciences@acdscience.va

PONTIFICIAE ACADEMIAE SCIENTIARVM ACTA
18

PATHS OF DISCOVERY

5-8 November 2004

EX AEDIBVS ACADEMICIS IN CIVITATE VATICANA

MMVI

The opinions expressed with absolute freedom during the presentation of the papers of this meeting, although published by the Academy, represent only the points of view of the participants and not those of the Academy.

ISBN 88-7761-088-3

The Holy Father Benedict XVI at Casina Pio IV.

The Holy Father John Paul II with the Plenary Session Participants, 8 November 2004.

Professor Nicola Cabibbo addresses the Holy Father John Paul II, 8 November 2004.

H.Em. Joseph Card. Ratzinger leaving Holy Mass concelebrated
for the Academicians at Montecassino Abbey, 7 November 2004.

Professor Werner Arber presents his paper, 6 November 2004.

Some Participants while Professo Carlo Rubbia presents his paper, 8 November 2004.

CONTENTS

PREFACE

It is a great honour for me to present this volume which contains the proceedings of the Plenary Session of the Pontifical Academy of Sciences on 'Paths of Discovery' which was held in November 2004. This volume demonstrates that the relativistic, nihilistic and fundamentalistic tendencies of some currents within modernity, which have been criticised by Benedict XVI and his predecessors with increasing force, have been matched by a new sense of realism and a welcome and progressive return of philosophical, ethical and theological questions. As this collection of papers observes, science today is experiencing a period of unforeseen and unpredictable development. With notable insight, Pius XI had envisaged this evolution of modern science when he renewed the Pontifical Academy of Sciences in 1936 and appointed as its members some of the most important scientists of the day.

The recent results of research into nuclear and sub-nuclear particles, directed towards analysing the structure of matter at a fundamental level, have been especially significant. This evolution, with its paths of discovery, which until half a century ago appeared unimaginable, is today in a state of continual expansion. The latest advances in astrophysics, which, indeed, are described in this volume, are particularly surprising. One need only think of the huge satellites that are in orbit around Venus and Saturn and are collecting material to be employed to compare these planets, which are of the same age as ours, with the planet earth. All these discoveries constitute a further confirmation of the analogical unity of physics and astrophysics. Similar observations may be made about biology, with the discovery of the basic micro- and macro-molecular structures of forms of life. Here it appears that mankind has entered a new era of knowledge.

As Pope John Paul II observes in the address that is published in this volume, 'the 'wonder' which sparked the earliest philosophical reflection on nature and which gave rise to science itself', and has marked the paths taken by science, has not diminished but increased with the new discoveries in physics, astrophysics and biology. This new world, which has been increasingly investigated by man, has given rise to an even greater 'wonder' at the universe which has extended the gap between our consciousness of our knowledge of nature and the reality itself of nature. To quote Shakespeare: 'There are more things in heaven and earth, Horatio, than are dreamt of in your philosophy'.[1] The strong connection that is being discovered between the primary forces of the cosmos and the apparently ultimate particles of matter attests to the fact that man, at least at his corporeal level, participates in the physical and biological bases of the totality of nature. And it is man once again who, immersing himself in this mystery of the created world, can expand the project of his being, as, indeed, was envisaged by Heraclitus with the Logos and by Aristotle who saw 'the intellect as capable of becoming and doing all things'.[2]

This new positive sense of the mystery of the Biblical Creation challenges man to provide a further philosophical and theological response, and also tells him that life on earth will end and that the 'Kingdom of God is near'.[3] In this way, as a result of the new paths of science – one of the principal protagonists of contemporary culture – religion and philosophy, which appeared to be superseded, have returned to the fore, as is demonstrated by the new discipline of bioethics and the increasing attention paid to the relationship between science and religion and between faith and reason. The Pontifical Academy of Sciences intends to return to these areas of controversy with the aim of fostering conditions for a new reconciliation and a peaceful synergy.

A beginning in this direction was indicated during the Holy Mass celebrated on the Sunday during the Plenary Session and presided over by Cardinal Ratzinger, a member of the Pontifical Academy. In his important homily on that occasion (which is published in this collection of papers), the now Benedict XVI, when referring to Holy Scripture and implicitly addressing the above-mentioned questions, observed that the 'reason why human life stretches beyond death' is man's 'being-in-relationship' with

[1] *Hamlet*, act 1 sc. 5.
[2] *De Anima*, 430 a 14 ff.
[3] *Lk* 21:31.

God: 'the communion of love with God is essential to love: this relationship never breaks up. If we are with Him then we are in the real, the indestructible life'. In conclusion, the Cardinal as he then was, observed that a human person in this relationship 'knows that he can and may risk his biological life, that he does not fall into void, because by taking this risk he takes hold of the true life'.

In this preface it is most fitting to thank His Holiness John Paul II, whose last audience to the Pontifical Academy was given during this Plenary Session, for everything that he did for this pontifical institution. During the twenty-seven years of his pontificate, John Paul II appointed a hundred and six Academicians, celebrated in 1986 the fiftieth anniversary of the refoundation of the Academy by Pius XI as well as the four hundredth anniversary, in 2003, of its foundation under Clement VIII, and was the moving spirit behind the restoration of the Casina Pio IV which was completed in the same year. John Paul II addressed the Academy more than thirty times on subjects that ranged from the frontiers of scientific knowledge to the role of science in ensuring the future of human life on the planet earth and meeting the needs of the poorest populations. This great Pope will also be remembered in the history of science for the solemn conclusion in 1992 to the Galileo case. This constituted a further contribution to this Pontiff's goal of reconciling science and faith, and formed a part of his involvement of science at the service of peace.

Perhaps one of the greatest gifts of John Paul II to the Pontifical Academy of Sciences was the revival of the tradition of appointing Cardinals as members of the Academy, starting with Joseph Ratzinger and Carlo Maria Martini and continuing with Georges Cottier, thus paving the way for the election of a member of the Academy as the successor of St Peter in the person of Benedict XVI, as had happened with Cardinal Eugenio Pacelli who became Pius XII. This election was of immense comfort to the Academicians at the time of the loss of John Paul II, an unforgettable patron and benefactor of the Pontifical Academy of Sciences.

I would also like to thank Prof. Werner Arber who originally suggested the subject for this Plenary Session and was its organiser at all stages. Words of gratitude should also be extended to President Nicola Cabibbo and the members of the Council of the Academy for their important contribution to the success of this Plenary Session. Lastly, all the participants must also be thanked for having honoured the Pontifical Academy at this meeting with their most valuable papers and illuminated discussions.

We all offer up our prayers that the pontificate of Benedict XVI which has begun with so many propitious signs of grace may meet with a renewed spirit of research and study within the Pontifical Academy, directed towards service to the Truth and the common good of the human family.

✠ Marcelo Sánchez Sorondo

INTRODUCTION TO THE SUBJECT OF THE PLENARY SESSION 'PATHS OF DISCOVERY'

WERNER ARBER

The membership of the Pontifical Academy of Sciences covers all the various fields of the natural sciences with some ramifications into the social sciences. For its plenary sessions the Academy intends to debate on a subject that encounters the wide interests of all its members. The topic 'Paths of Discovery' chosen for the plenary session of November 2004 corresponds well to this request. Generally speaking, discoveries are the fruits of intensive research. They bring about stepwise, smaller or larger additions to – or modifications of – the actual scientific knowledge base. Scientific discoveries represent innovations, they are the results of creative acts. Scientific knowledge is an important basis for many kinds of applications of the acquired knowledge.

First, new knowledge can open novel research strategies for both fundamental and applied scientific research. Often, it happens that a new research strategy is the source of a whole series of new discoveries. One might conclude that an appropriate choice of research strategies is a good guarantee for fruitful discoveries.

Second, scientific knowledge offers many types of practical applications, often of technological use for the benefit of mankind. In general, these applications require additional research, so-called applied research. Discoveries often depend on the alertness of the investigators. They can very well occur in steps towards a practical application of previously acquired scientific knowledge; they are not limited to fundamental research.

Third, scientific knowledge has a strong and lasting impact on the world view of human beings. This principally represents philosophical values and it forms a fundamental basis of the orientational and dispositive knowledge that serves the civil society in policy decisions, some of which function as regulators for practical, technological applications of scientific knowledge.

These considerations reflect the cultural values of scientific knowledge that were the subject of the plenary session of the Pontifical Academy of Sciences in November 2002.[1] It was thus consistent to have a closer view this time of the various paths in research activities that pave the way to obtain new insights into the natural reality and thus to make scientific discoveries.

In the invitation to the members of the Academy to attend the plenary session in November 2004, the Council of the Pontifical Academy of Sciences characterized the aim of the debate on 'Paths of Discovery' with the following text:

> 'Discoveries are at the basis of new knowledge. There is no single recipe leading to discovery; there are a multitude of paths. Some discoveries are made upon verification or falsification of a theory. The discovery can then also give rise to a refinement of the theory, paving the way for further experimental approaches. Quite frequently, a discovery is made while the researcher is looking for something else. His scientific mind and intuition may thereby direct his attention to the unexpected. This source of discovery is generally called *serendipity*. Novel methodology and research strategies can open the doors to many new discoveries, which can then have an impact on individual researchers or, in other cases, on wider research programmes involving many researchers. Think of major scientific endeavors such as CERN depending on a large experimental set-up. However, in this case too, progress will depend in part on the activities of individual scientists and their aptitude to design and interpret experiments. The same is true for almost any research, whether it is carried out by a single person or in the context of a larger research program. More generally, the debate may also address the role played by dogma and widely anchored textbook knowledge in scientific progress, given their frequent influence on the interpretation of observations and experimental data.
>
> The aim of a debate on 'Paths of Discovery' is to collect a number of case studies largely based on the personal experience of the participants and their scientific environments. However, the debate may also refer to some cases of discovery that have already been

[1] *The Cultural Values of Science*, The Pontifical Academy of Sciences, Scripta Varia *105*, Vatican City, 2003.

documented by the history of science. The resulting documentation is expected to illustrate the range of paths leading to discovery. Such documentation may be useful in political planning for scientific investigations and may represent a welcome contribution to the literature on the principles of the scientific approach'.

The collection of contributions made by members of the Academy gives a wide spectrum of potential access to scientific discoveries, to scientific innovation. This collection should, however, not be expected to be a systematic, complete presentation of the many paths leading to discoveries. Rather, this publication should be seen as a collection of reports on specific cases in which a member of the Academy was personally involved either in the research leading to a discovery or in studies addressing paths of discovery from the point of view of history of science. We expect that this documentation will find interested readers within the community of scientists, among the historians of science, among people active in science politics and planning and also among the general public.

PROGRAMME

THURSDAY 4 NOVEMBER 2004

16.00 *Council Meeting*

FRIDAY 5 NOVEMBER 2004

9:10 *Word of Welcome*: Prof. Nicola CABIBBO, President of the Academy
9:15 *The Subject of the Meeting*: Prof. Werner ARBER, Coordinator of the Meeting
9:20 *Commemorations of Deceased Academicians*
Sune BERGSTRÖM (15.VIII.04) by Prof. N.M. LE DOUARIN
Nicola DALLAPORTA (23.X.03) by Prof. N. CABIBBO
Renato DARDOZZI (3.VI.03) by Prof. N. CABIBBO
Paul A.J. JANSSEN (11.XI.03) by Prof. C. DE DUVE
Stanisław ŁOJASIEWICZ (13.XI.02) by Prof. P. GERMAIN
Manuel LORA TAMAYO (22.VIII.02) by Prof. A. GARCÍA BELLIDO
Thomas R. ODHIAMBO (27.V.03) by Prof. M.G.K. MENON
George PORTER (31.VIII.02) by Prof. R. HIDE
Marcel ROCHE (3.V.03) by Prof. C. PAVAN
10:20 *Self-Presentations of the New Members*
Prof. Suzanne CORY
Prof. Antonio GARCÍA BELLIDO
Prof. William D. PHILLIPS
Prof. Yves QUÉRÉ
Prof. Veerabhadran RAMANATHAN
10:10 Coffee Break
11:25 Session One
Chairperson: Prof. W. ARBER
Speaker: Prof. J. MITTELSTRASS
Different Types of Discovery – Lessons from the History of Science
11:45 Discussion
11:55 Speaker: Prof. G.V. COYNE
Discovery in the New Cosmology of Copernicus, Kepler and Galileo

12:20 *Speaker*: Prof. P.J. LÉNA
The Case of the Extrasolar Planets
13:00 General Discussion
13:15 Lunch at the Academy
15:00 Session Two
Chairperson: Prof. G.V. COYNE
Speaker: Prof. R. MURADIAN
Going from Quarks to Galaxies: Two Findings
15:30 Discussion
15:35 Speaker: Prof. B.M. COLOMBO
Paths to Discovery: Personal Experiences in a Social Science
15:45 Discussion
16:00 Speaker: Prof. S.L. JAKI
On a Discovery about Godel's Incompleteness Theorem
16:15 General Discussion
16:30 Coffee Break
16:45 *Presentation of the Pius XI Medal*
Awarded to Prof. Laure SAINT-RAYMOND
17:05 Session Three
Chairperson: Prof. P. GERMAIN
Speaker: Prof. M. MOSHINSKY
Transient Phenomena in Quantum Mechanics: Diffraction in Time
17:30 Speaker: Prof. C. COHEN-TANNOUDJI
Optical Methods – A Simple Way to Interrogate and to Manipulate Atoms
17:50 Discussion
18:00 Speaker: Prof. A.M. BATTRO
Microdiscoveries: a Case Study of Creative Paths and Academic Networks in Science
18:30 Discussion
18:40 Speaker: Prof. J.-M. MALDAMÉ
Comment la réflexion sur l'inspiration de la Bible éclaire l'invention en matière de science
19:15 General Discussion
19:45 Dinner at Domus Sanctae Marthae

SATURDAY 6 NOVEMBER 2004

9:10 Session Four
Chairperson: Prof. P.J. CRUTZEN
Speaker: Prof. W.E. THIRRING
Instabilities of Gravitation Dominated Systems

9:25 Discussion
9:30 Speaker: Prof. M.J. Rees
From Confusion towards Consensus in Cosmology
10:00 Discussion
10:20 Speaker: Prof. A. Zichichi
A Personal Experience of Unexpected Discoveries
10:45 Discussion
11:00 Coffee Break
11:30 Speaker: Prof. M. Heller
Discovering the World Structure as a Goal of Physics
11:45 Discussion
12:00 General Discussion
12:15 Session Five
Chairperson: Prof. P.J. Léna
Speaker: Prof. M. Sela
Paths of Discovery of a Therapeutic Vaccine
12:30 Discussion
12:45 Lunch at the Academy
14:50 Speaker: Prof. A. Szczeklik
Aspirin and Eicosanoid Pathways: The Discoveries that Changed Contemporary Medicine
15:15 Discussion
15:25 Speaker: Prof. M.G.K. Menon
The Role of Innovation, Interdisciplinarity and Phenomenology as Components of Creativity in Opening New Windows
16:00 Discussion
16:15 Speaker: Prof. P.J. Crutzen
The 'Anthropocene', the Present Era Influenced by Human Activity
16:40 Discussion
16:50 Speaker: Prof. R.J. White
Discovering Pathways to Brain Cooling
17:15 Discussion
17:20 Coffee Break
17:40 Session Six
Chairperson: Prof. A.M. Battro
Speaker: Prof. W. Arber
The Impact of Microbial Genetics on the Development of Genomics and Biotechnology
18:20 Discussion

18:30 Speaker: Prof. C. Pavan
Endosymbiotic Bacteria Associated with Plant Seeds and Birds' Eggs
18:50 Discussion
19:00 Speaker: Prof. R. Vicuña
Microbial Biodiversity: a New Voyage of Discovery
19:15 Discussion
20:00 Dinner at Domus Sanctae Marthae

Sunday 7 November 2004

Trip to Montecassino Abbey 60th Anniversary of the Abbey's Destruction and Reconstruction
7:00 Bus leaves Domus Sanctae Marthae
10:15 Arrival in Montecassino and Greeting by Abbot Monsignor Bernardo D'Onorio
10:30 *Abbey Mass with Gregorian Chants, celebrated by His Eminence Cardinal Joseph Ratzinger*
11:30 Guided Visit to the Abbey
12:30 Social Lunch at the Abbey
15:00 Bus leaves for Domus Sanctae Marthae

Monday 8 November 2004

9:10 Session Seven
Chairperson: Prof. V.I. Keilis-Borok
Speaker: Prof. R. Hide
Geomagnetism, 'Vacillation', Atmospheric Predictability and 'Deterministic Chaos'
9:45 Discussion
10:00 Speaker: Prof. C. Rubbia
Energy and Poverty
10:30 Discussion
10:45 Coffee Break
11:30 *Papal Audience and Photograph with the Holy Father*
13:00 Lunch at the Academy
15:00 Closed Session for Academicians
19:00 Dinner at Domus Sanctae Marthae

LIST OF PARTICIPANTS

Academicians

Cabibbo Prof. Nicola (President)
Università degli Studi di Roma 'La Sapienza'
Dipartimento di Fisica
Piazzale Aldo Moro, 5
I-00185 Roma (Italy)

Cottier Card. Georges Marie Martin, O.P.
Palazzo Apostolico
V-00120 Città del Vaticano

Martini Card. Carlo Maria, S.J.
Santuario di Galloro
Via Appia, 54
I-00040 Galloro, RM (Italy)

Ratzinger Card. Joseph
Congregazione per la Dottrina della Fede
Palazzo del S. Uffizio
V-00120 Città del Vaticano

Sánchez Sorondo H.E. Msgr. Marcelo (Chancellor)
Pontificia Accademia delle Scienze
Casina Pio IV
V-00120 Città del Vaticano

Arber Prof. Werner
University of Basel, Biozentrum
Department of Microbiology
Klingelbergstrasse 70
CH-4056 Basel (Switzerland)

Battro Prof. Antonio M.
Battro & Denham, Consultores
Billinghurst 2574 Piso 1A
C1425DTZ Buenos Aires (Argentina)

Cohen-Tannoudji Prof. Claude
Collège de France – École Normale Supérieure
Département de Physique – Laboratoire Kastler Brossel
24, rue Lhomond
F-75231 Paris Cedex 05 (France)

Colombo Prof. Bernardo Maria
Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Via C. Battisti, 241
I-35121 Padova (Italy)

Cory Prof. Suzanne
The Walter and Eliza Hall Institute of Medical Research
1G Royal Parade
Parkville
Victoria 3050 (Australia)

Coyne Prof. George V., S.J.
Specola Vaticana
V-00120 Città del Vaticano

Crutzen Prof. Paul Josef
Max-Planck-Institute for Chemistry, Department of Atmospheric Chemistry
P.O. Box 3060
D-55020 Mainz (Federal Republic of Germany)

García Bellido Prof. Antonio
Universidad Autónoma de Madrid
Centro de Biología Molecular ‘Severo Ochoa’
Laboratorio de Genética del Desarrollo
Cantoblanco
28049 Madrid (Spain)

Germain Prof. Paul Marie
Académie des Sciences
23, quai de Conti
F-75006 Paris (France)

Heller Prof. Michael
Pontifical Academy of Theology
Faculty of Philosophy
ul. Franciszkańska, 1
PL-31-004 Kraków (Poland)

Hide Prof. Raymond
Department of Mathematics
Imperial College
180 Queen's Gate
London SW 7 2BZ (United Kingdom)
Mailing address:
17, Clinton Avenue
East Molesey, Surrey KT8 0HS (United Kingdom)

Jaki Prof. Stanley L., O.S.B.
Seton Hall University
South Orange, NJ, 07079 (USA)
Mailing address:
P.O. Box 167
Princeton, NJ 08542 (USA)

Keilis-Borok Prof. Vladimir Isaakovich
University of California, Los Angeles
Institute of Geophysics and Planetary Physics
3845 Slichter Hall, Box 951567
Los Angeles, CA 90095-1567 (USA)
Int. Institute of Earthquake Prediction Theory
and Mathematical Geophysics
Warshavskoye sh. 79, Kor 2
Moscow 113556 (Russia)

Le Douarin Chauvac Prof. Nicole Marthe
Académie des Sciences
23, quai de Conti
F-75006 Paris (France)
C.N.R.S. – Collège de France
Institut d'Embriologie Cellulaire et Moléculaire – UPR 9064
49bis, avenue de la Belle Gabrielle
F-94736 Nogent-sur-Marne Cedex (France)

Léna Prof. Pierre Jean
Université Paris VII Denis Diderot – Observatoire de Paris
Département de Recherche Spatiale
Unité de recherche associée au CNRS # 8632
Place Jules-Janssen
F-92195 Meudon (France)

Levi-Montalcini Prof. Rita
Institute of Neurobiology, CNR
Viale Marx, 15
I-00137 Roma (Italy)

Maldamé Prof. Jean-Michel, O.P.
Institut Catholique de Toulouse
31, rue de la Fonderie
F-31068 Toulouse Cedex (France)

Menon Prof. Mambillikalathil Govind Kumar
K-5 (Rear), Hauz Khas Enclave
New Delhi 110016 (India)

Mittelstrass Prof. Jürgen
Universität Konstanz
Fachbereich Philophie und Zentrum Philosophie
und Wissenschaftstheorie
D-78457 Konstanz (Federal Republic of Germany)

Moshinsky Prof. Marcos
Universidad Nacional Autónoma de México
Instituto de Física
Apdo. Postal 20-364, Delegación Alvaro Obregon
01000 México, D.F. (México)

Muradian Prof. Rudolf
Byurakan Astrophysical Observatory
AM-378433 Byurakan (Armenia)
Mailing address:
Universidade Estadual de Santa Cruz
Departamento de Ciências Exatas e Tecnológicas
Rodovia Ilhéus/Itabuna, km 16
45650-000 Ilhéus, Bahia (Brazil)

Pavan Prof. Crodowaldo
Universidade de São Paulo, USP
Instituto de Ciências Biomédicas, Lab. de Microbiologia
São Paulo, S.P. 05389-970 (Brazil)

Phillips Prof. William Daniel
National Institute of Standards and Technology
Physics Laboratory – Atomic Physics Division
100 Bureau Drive, Stop 8424
Blog. 216, Rm. B137
Gaithersburg, MD 20899 (USA)

Quéré Prof. Yves
Académie des Sciences
Délégation aux Relations internationales
23, quai de Conti
F-75006 Paris (France)

Ramanathan Prof. Veerabhadran (Ram)
University of California, San Diego
Scripps Institution of Oceanography
Center for Atmospheric Sciences
9500 Gilman Drive, MC 0221
La Jolla, CA 92093-0221 (USA)

Rees Prof. Martin John
University of Cambridge
Institute of Astronomy
Madingley Road
Cambridge CB3 0HA (United Kingdom)

Rubbia Prof. Carlo
Ente per le Nuove Tecnologie, l'Energia e l'Ambiente (ENEA)
Lungotevere Thaon di Revel, 76
I-00196 Roma (Italy)
European Organization for Particle Physics (CERN)
CH-1211 Geneva 23 (Switzerland)

Sela Prof. Michael
The Weizmann Institute of Science
Department of Immunology
P.O. Box 26, Rehovot 76100 (Israel)

Szczeklik Prof. Andrzej
Jagiellonian University School of Medicine
Department of Medicine
ul. Skawińska 8
PL-31-066 Kraków (Poland)

Thirring Prof. Walter E.
Universität Wien
Institut für Theoretische Physik
Boltzmanngasse 5
A-1090 Vienna (Austria)

Tuppy Prof. Hans
University of Vienna Institute of Medical Biochemistry
Dr. Bohr-Gasse 9/3. Stock
A-1030 Vienna (Austria)

Vicuña Prof. Rafael
Pontificia Universidad Católica de Chile
Facultad de Ciencias Biológicas
Departamento de Genética Molecular y Microbiología
Casilla 114-D, Santiago (Chile)

White Prof. Robert Joseph
Case Western Reserve University Medical School
MetroHealth Medical Center
Division of Neurological Surgey and Brain Research Laboratory
2500 MetroHealth Drive
Cleveland, OH 44109-1998 (USA)

Zichichi Prof. Antonino
'Ettore Majorana' Foundation and Centre for Scientific Culture
Via Guarnotta, 26
I-91016 Erice (Italy)

Pius XI Medal

Saint-Raymond Prof. Laure
Laboratoire Jacques-Louis Lions
Université Paris VI Pierre et Marie Curie
175, rue du Chevaleret
F-75013 Paris (France)

Observers

Kashina Dr. Anna S.
University of Pennsylvania
School of Veterinary Medicine
Department of Animal Biology
3800 Spruce Street, 143 Rosenthal
Philadelphia, PA 19104 (USA)

Tanzella-Nitti Rev. Prof. Giuseppe
Library
Pontificia Università della Santa Croce
Via dei Farnesi, 82
I-00186 Roma (Italy)

ADDRESS OF THE PRESIDENT TO THE HOLY FATHER

Holy Father,

Let me first express our gratitude for the precious gift you are bestowing on the Pontifical Academy of Sciences with this audience.

One year ago, in November 2003 we celebrated the fourth anniversary of the foundation of the Lincei from which our Academy descends. This meeting is also a special one, the first Plenary Session of the Academy in its fifth century, and we decided to devote it to a discussion of the very foundation of scientific research, the creative power of the human mind, and the way in which the admiration and contemplation of the beauty of the natural world gives rise to often unexpected scientific discoveries.

During this year the Academy devoted two workshops to subjects that are of the utmost importance for human life. The first is water, 'Nostra sorella acqua' after St Francis, whose presence or absence can determine the birth or downfall of entire civilizations. The second meeting, held last week – many of the participants are here today – was devoted to a problem which elicits the highest concerns in the scientific world: the possible impacts of the expected climate changes on human health.

Also during this session we welcomed five new members of the Academy, and I have the honour of presenting them to Your Santity:

Prof. Yves Quéré French Physicist
Prof. Garcia Bellido, Spanish Biologist, Developmental Genetics,
Prof. Suzanne Cory, Australian Molecular Biologist
Prof. William D. Phillips, American Physicist, Nobel Prize for Physics 1997
Prof. Veerabhadran Ramanathan, Indian Earth Scientist.
Prof. Fotis Kafatos, Greek Molecular Biologist
Prof. Tsung-Dao Lee, Chinese Physicist

We thank you again, Holy Father, for receiving us today!

Nicola Cabibbo

ADDRESS OF JOHN PAUL II TO THE PARTICIPANTS IN THE PLENARY SESSION OF THE PONTIFICAL ACADEMY OF SCIENCES

Monday, 8 November 2004

Ladies and Gentlemen,
Dear Friends,

1. It is with particular pleasure that I greet the distinguished members of the *Pontifical Academy of Sciences*. I thank your President, Professor Nicola Cabibbo, for the kind message of greetings and good wishes which he has offered me in your name.

The meetings of the Academy have always been an occasion of mutual enrichment and, in some cases, have resulted in studies of significant interest to the Church and the world of culture. These initiatives have contributed to a more fruitful dialogue between the Church and the scientific community. I trust that they will lead to an ever deeper investigation of the truths of science and the truths of faith, truths which ultimately converge in that one Truth which believers acknowledge in its fullness in the face of Jesus Christ.

2. This year's plenary session, devoted to science and creativity, raises important questions deeply connected with the spiritual dimension of man. Through culture and creative activity, human beings have the capacity to transcend material reality and to 'humanize' the world around us. Revelation teaches that men and women are created in the 'image and likeness of God'[1] and thus possessed of a special dignity which enables them, by the work of their hands, to reflect God's own creative activity.[2] In a real way, they are meant to be 'co-creators' with God, using their knowledge and

[1] Cf. *Gen* 1:26.
[2] Cf. *Laborem Exercens*, 4.

skill to shape a cosmos in which the divine plan constantly moves towards fulfilment.[3] This human creativity finds privileged expression in the pursuit of knowledge and scientific research. As a spiritual reality, such creativity must be responsibly exercised; it demands respect for the natural order and, above all, for the nature of each human being, inasmuch as man is its subject and end.

The creativity which inspires scientific progress is seen especially in the capacity to confront and solve ever new issues and problems, many of which have planetary repercussions. Men and women of science are challenged to put this creativity more and more at the service of the human family, by working to improve the quality of life on our planet and by promoting an integral development of the human person, both materially and spiritually. If scientific creativity is to benefit authentic human progress, it must remain detached from every form of financial or ideological conditioning, so that it can be devoted solely to the dispassionate search for truth and the disinterested service of humanity. Creativity and new discoveries ought to bring both the scientific community and the world's peoples together, in a climate of cooperation which values the generous sharing of knowledge over competitivity and individual interests.

3. The theme of your meeting invites renewed reflection on the 'paths of discovery'. There is in fact a profound inner logic to the process of discovery. Scientists approach nature with a conviction that they confront a reality which they have not created but received, a reality which slowly reveals itself to their patient questioning. They sense – often only implicitly – that nature contains a Logos which invites dialogue. The scientist seeks to ask the right questions of nature, while at the same time maintaining an attitude of humble receptivity and even of contemplation in its regard. The 'wonder' which sparked the earliest philosophical reflection on nature and which gave rise to science itself, has in no way been diminished by new discoveries; indeed, it constantly increases and often inspires awe at the distance which separates our knowledge of creation from the fullness of its mystery and grandeur.

Contemporary scientists, faced with the explosion of new knowledge and discoveries, frequently feel that they are standing before a vast and infinite horizon. Indeed, the inexhaustible bounty of nature, with its promise of ever new discoveries, can be seen as pointing beyond itself to

[3] Cf. *Gaudium et Spes*, 34.

the Creator who has given it to us as a gift whose secrets remain to be explored. In attempting to understand this gift and to use it wisely and well, science constantly encounters a reality which human beings 'find'. In every phase of scientific discovery, nature stands as something 'given'. For this reason, creativity and progress along the paths of discovery, as in all other human endeavours, are ultimately to be understood against the backdrop of the mystery of creation itself.[4]

4. Dear members of the Academy, once again this year I offer my prayerful good wishes for your work on behalf of the advancement of knowledge and the benefit of the human family. May these days of reflection and discussion be a source of spiritual enrichment for all of you. Despite the uncertainties and the labour which every attempt to interpret reality entails – not only in the sciences, but also in philosophy and theology – the paths of discovery are always paths towards truth. And every seeker after truth, whether aware of it or not, is following a path which ultimately leads to God, who is Truth itself.[5] May your patient and humble dialogue with the world of nature bear fruit in ever new discoveries and in a reverent appreciation of its untold marvels. Upon you and your families I cordially invoke God's blessings of wisdom, joy and peace.

[4] Cf. *Laborem Exercens*, 12.
[5] Cf. *Fides et Ratio*, 16, 28.

COMMEMORATIONS OF DECEASED ACADEMICIANS

SUNE BERGSTRÖM († 15.VIII.04)

The Swedish biochemist Sune Bergström was awarded the Nobel Prize for Medicine and Physiology in 1982 together with Bengt I. Samuelson and John R. Vane, for 'their isolation, identification and analysis of *prostaglandins*'. Prostaglandins are biochemical compounds that influence blood pressure, body temperature, allergic and inflammatory reactions and other physiological phenomena in mammals.

Sune Bergström was the first to demonstrate that prostaglandins were produced in different forms, to determine their chemical nature and to decipher the biosynthetic pathways producing them. For this pioneer work in the identification of what can be considered as new hormones, produced by a number of cells in the body, Bergström is often called the 'father' of prostaglandins. There compounds can have both positive and negative effects on the body. His discoveries opened a very active avenue of investigations on the metabolism of unsaturated fatty acids and led to the development of new drugs that counteract some of the effects of prostaglandins.

Sune Bergström was born in 1916 in Stockholm. He received his scientific education at the Karolinska Institute in Stockholm where he was awarded doctoral degrees in Medicine and Biochemistry in 1944. He held research fellowships at Columbia University and at the University of Basel. He then returned to Sweden and became Professor of Chemistry at the University of Lund. In 1958, Bergström returned to the Karolinska Institute, and became Dean of the Medical Faculty in 1963 and Rector in 1969.

After retiring from teaching in 1981, he continued to conduct research. He was chairman of the Nobel Foundation (1975-1987) and chairman of Medical Research at the World Health Organization (from 1977 to 1982).

Sune Bergström was deeply involved in health problems in developing countries in which prostaglandins and related drugs can be used to relieve suffering from tropical diseases, nutrition and birth control problems.

Sune Bergström passed away on August 15, 2004 and the world lost an outstanding scientist as well as an eminent humanist.

Nicole M. Le Douarin

NICOLA DALLAPORTA († 23.X.03)

Nicola Dallaporta died in October 2003. He had been a respected and active member of our Academy since 1989. A graduate of Bologna University, he taught at the Universities of Catania, Turin and Padua, where he was nominated full professor in 1947. Dallaporta was one of the leaders of the reconstruction, in the post-war years, of the Italian school of Theoretical Physics. His early work was of very high quality, and touched on the composition of cosmic rays, on the nature of strange particles and their weak interactions, on the properties of hypernuclei, which are new types of atomic nuclei where one of the protons or neutrons is substituted with a strange particle, such as a 'Lambda' hyperon, and on the symmetry properties of elementary particles. His work had wide resonance, and has left its mark in the development of elementary particles. We find it cited in history books, for example in 'Inward Bound' by Abraham Pais.

In the sixties Dallaporta gave a crucial contribution to the development of Italian Physics, with the foundation of the first group of Theoretical Astrophysics in Padua. Italian astronomy was then a very secluded discipline, centered in a handful of observatories, often endowed with antiquated instrumentation. In opening up the new field of astrophysics, Dallaporta brought to astronomy the team work style and international flavour well established in the high energy physics community, and this soon bore fruit within a number of universities. His astrophysics work, which he pursued into his later years, was also of consistently high quality and widely appreciated. It ranged from stellar evolution to cosmology and the dynamical evolution of galaxies.

Dallaporta was not only a great scientist, but a person of great kindness and humanity. The writer Primo Levi, author of *If This Is A Man*, fondly remembered his close friendship with Dallaporta, how he helped him in his studies in Turin, and after the war tried to help him again in his attempts to recover from the horror of the nazi concentration camps. Dallaporta was also a man of profound religious convictions, who attained an admirable harmony between religion and science. In the very last years of his life Dallaporta took an active part in the discussion of the anthropic principle, to which he contributed his original viewpoint.

I will not mention the many honours Dallaporta received, which you can find listed in our Yearbook, but only express my deep sense of loss for his departure.

Nicola Cabibbo

RENATO DARDOZZI († 3.VI.03)

Renato Dardozzi directed the activities of our Academy as Director of the Chancery from 1985, and as Chancellor in 1996. In 1997 he was appointed Honorary Member of the Academy. For the new members, who do not remember his untiring activity in the promotion of our Academy, I will recall some aspects of his life. He started his career as an engineer specialised in the field of telecommunications, which was then on the verge of the tumultuous growth we are now witnessing. He found employment with the Italian telephone company, and through his brilliant intelligence and personal integrity rose through the ranks, finally becoming the Director General of STET, the Italian Telecommunications Holding, now privatised as Telecom Italia.

During his successful career he felt ever more attracted to a religious life, and enrolled in a Theology and Philosophy course at the Gregorian University in Rome. After graduating in 1973 he took Holy Orders. His managerial prowess had not however passed unnoticed in the Vatican, and the following year he was called to work for the Secretary of State. So his hope for a life devoted to thought and philosophy had to give way to a renewed manegerial activity at the highest level. Apart from leading the activities of the Academy, Dardozzi was many times called to very delicate tasks, from that of reviewing the management of the Vatican Bank, which had attracted widespread criticism, to that of restoring the scientific and medical excellence of the 'Bambin Gesù' hospital, which is now one of the most highly regarded pediatric hospitals in Italy.

When our sister institution, the Pontifical Academy of Social Sciences, was founded, Dardozzi was also nominated Chancellor of that Academy, and had an important role in the startup of that institution.

To the tireless activity of Msgr. Dardozzi must go a good part of the merit for the high level of the many meetings which took place under his tenure, and attracted a great deal of attention at the international level. Among them 'Population and Resources', 'The Epoch of Galaxy Formation', 'Chemical Hazards in Developing Countries', 'Legal and Ethical Problems of the Human Genome Project', 'The Origin and Early Evolution of Life'. Dardozzi initiated the publication of the proceedings of some of our meetings with prestigious publishers such as the Cambridge or Oxford University Presses, and this idea has been followed to the present day. I must also recall that Dardozzi had an important role in organizing the work for the review of the Galilei process, required by John Paul II, which was crowned by the solemn declaration of the Holy Father in 1992.

Dardozzi was an Engineer, a title he always insisted on, even in his priestly life, and always maintained a profound interest for the philosophical implications of the new scientific discoveries. He was also a man of great charity, which was manifest in his gentle attention to the personal problems of our employees, but also in his private life, which was dedicated to the care of a girl, Federica, affected by serious handicaps, whom he had adopted as a daughter. I really hope we can convince one of his many friends to write the story of the very interesting life of this unique person who graced the Academy with his dedicated work.

Nicola Cabibbo

PAUL A.J. JANSSEN († 11.XI.03)

Paul Adriaan Jan Janssen was born on September 12, 1926, in Turnhout, Belgium. He passed away on November 11, 2003, in Rome, while attending the celebration of the 400th anniversary of the Pontifical Academy of Sciences, of which he was a member since 1990.

A descendant of a farmer family from the Kempen region, Paul Janssen's father, Dr Constant Janssen, had become a successful general practitioner when, in 1933, he founded a small pharmaceutical company dealing mostly with the marketing of products from a Hungarian concern. After completing his humanities, in 1943, his son Paul decided to follow in his father's footsteps, but in a more creative and ambitious way. He would develop his own pharmaceuticals.

To prepare for his chosen goal, the young Paul Janssen decided to study medicine, which he did in three Belgian universities, first at the Facultés Notre-Dame de la Paix in Namur (1943-1945), then at the Catholic University of Louvain (1945-1949) and finally at the University of Ghent, where he graduated in 1951 and continued as a part-time assistant in the pharmacology laboratory of Nobel Laureate Corneille Heymans, with whom he remained associated until 1956. Short stays in the United States and in several European countries completed his training. But he soon cut his academic ties. The desire to be on his own was too strong.

As early as 1953, the young doctor entered the family business and embarked on research with a skeleton staff. Only five years later, five new compounds had already been synthesized and introduced into clinical use. By 1961, the company already employed 400 people and had attracted the attention of the American industrial giant Johnson & Johnson. The resulting merger provided Paul Janssen with security, while leaving him an entirely free hand. Under his leadership, the company, which was named 'Janssen Pharmaceutica' in 1964, developed into a world-wide consortium, which now employs more than 4,000 people in Belgium alone and has affiliates in a large number of countries, including the United States, Japan, and China, where it was implanted as early as 1985.

The phenomenal success of this enterprise was due to the unique qualities of Paul Janssen, who was, at the same time, a true chemical genius, a man of exceptional vision, a remarkable entrepreneur and an effective and respected leader. Paul Janssen was the holder of more than 100 patents and the author of more than 800 publications. His outstanding merits have been recognized by many awards, membership in a number of academies, including the Pontifical Academy of Sciences, 22 honorary doctorates, and many other national and foreign distinctions. He was elevated to the rank of baron by H.M. King Baudouin in 1990.

Paul Janssen was the loving husband of Dora Arts, whom he married in 1957, and the proud father of 5 children and grandfather of 13 grandchildren. In spite of his tremendous scientific and financial achievements, he was modest and tolerant, deeply devoted to his family. He appreciated art, supporting his wife in her collection of pre-Colombian artefacts, and loved music, being himself an accomplished pianist. He leaves the souvenir of a true humanist.

Christian de Duve

Stanisław Łojasiewicz († 13.XI.02)

Professor Stanisław Łojasiewicz was born on October 9, 1926 in Warsaw. He completed his mathematical studies in 1945-47 at the Jagiellonian University in Cracow. He was a student of Professor Tadeusz Wasewski and dealt at that time with differential equations, mainly ordinary, studying in particular asymptotic effects. He defended his PhD thesis in 1950.

He then turned his interests towards distribution theory, the systematic description of which had just been created by Laurent Schwartz. During a stay in Paris in 1957, he achieved great success solving the problem of division of distributions by analytic functions, a problem posed by Schwartz, and he published the result in a note of Comptes rendus de l'Académie des Sciences in 1958. Such a result finds many applications in the theory of partial differential equations and analytic geometry. The analysis of the method led Stanisław Łojasiewicz to create a new geometry, i.e. semianalytic geometry and also to initiate a generalization now called subanalytic geometry. These two new geometries became a useful new tool in many branches of analysis and control theory.

In 1956-60, Stanisław Łojasiewicz visited many universities, in Kingston, in Chicago, in Berkeley and also in Princeton at the Institute of Advanced Studies. In 1962, he obtained a professorship at the Jagiellonian University. He spent the year 1964-65 in Paris. In 1967-68, during his stay at the Institut des Hautes Etudes Scientifiques, he obtained a beautiful proof of the Malgrange-Mather Preparation Theorem. In 1970, he was invited to deliver a plenary lecture about semianalytic geometry at the International Congress of Mathematicians in Nice.

The intense scientific and didactic activity of Professor Łojasiewicz attracted many students and gave rise to a mathematical school, representatives of which are not only in Cracow, but also in many places in Poland and many centres abroad in France, Italy, Spain, Germany.

He was elected full member of the Polish Academy of Sciences in 1980 and to our 'Pontificia Academia Scientiarvm' in 1983. He was a member of our Council from 1989 to 1992.

He died on November 14, 2002 from a heart attack during his trip home to Cracow after attending our Plenary Session.

Paul Germain

Manuel Lora Tamayo († 22.VIII.02)

Professor Lora Tamayo died last August 2002 in Madrid. He was member of this Academia Pontificia since 1964. He was born in 1904 in Jerez de la Frontera in Andalucía (Spain). He studied in the Central University of Madrid where he obtained his PhD in Chemistry in 1930. He specialized in Biological Chemistry in Strasburg (France) and obtained his first University chair in 1933 in Sevilla and later in Madrid. He founded and intellectually guided a growing body of scientists thereafter, one of the most relevant schools of the Spanish biochemists. His work in organic chemistry and pharmacology is of a wide scope, dealing with subjects as distinct as sugar metabolism, organic condensations, phosphates with antituberculose potentialities and many more.

In addition to his scientific interests he was an outstanding leader in the organization of Science in the post-civil-war in Spain. He was

President of the Spanish Research Council, Minister of Education and Science (this second title was incorporated by him) and President of the Royal Spanish Academy of Science. He was member of several foreign and national Academies and received honorary degrees from many universities. His death was profoundly moving for the entire Spanish scientific community.

Antonio García Bellido

THOMAS R. ODHIAMBO († 27.V.03)

Thomas Odhiambo was the most influential and internationally known African scientist ever. He was born in Mombasa, Kenya on 4 February 1931 and died on 26 May 2003 in Nairobi, Kenya, at the age of 72.

Thomas Odhiambo came from a very poor background, being one of 10 children of a telegraph clerk. He was educated by missionaries and perhaps one of the important things that he acquired from them was a missionary zeal for service to humanity. Those of us who had the privilege of listening to him at the 2002 Plenary Session of our Academy will remember the way he expounded with missionary zeal that day.

Thomas Odhiambo took his degree in biology from Makerere College in Kampala, Uganda. Even before that, he always had a deep interest in science and nature; he used to study, for example, wasps, which most people regard as a nuisance because they sting; he was interested in their complicated social behavior. After he finished his degree, he worked for four years in the Ministry of Agriculture of the Uganda government, and with that practical experience went to Cambridge (UK) where he did his PhD in the area of insect physiology. He returned to the University of Nairobi to join the Zoology Department, where he created the first Department of Entomology.

In 1967, he wrote an important article in *Science* which defined his philosophy in life. It was entitled 'East Africa: Science for Development'. He had already by then got deeply interested in the problems of his country; in particular, he was moved by the plight of farmers fighting pests on their lands. Insecticides killed everything, good and bad, harmed the environment and were expensive. Did science have an alternative? He proposed the setting up of the International Centre for Insect Physiology and Entomology (ICIPE), in Nairobi, which has since become world famous. Here he analyzed traditional methods of farming in terms of the underlying science and improving on it made it available to farmers at low cost. In his article in *Science* he had also talked about capacity-building, the need for excellence, the need for relevance in scientific research and education, by working on real-life problems; and thus from his deep interest in wasps, his training in biology, particularly insect physiology, his commitment to his people and their plight, he moved on to the use of science in the service of humanity.

I remember well his election to this Academy in 1981, for that year Abdus Salam, an outstanding scientist and a great exponent of science for development and I were also elected. Soon thereafter, over a breakfast meeting, where all the Academians from the developing countries got together, including the then President, Carlos Chagas, we decided to set up a Third World Academy of Sciences – an idea discussed the previous night by Abdus Salam and myself. The Pontifical Academy of Sciences thus gave birth to the Third World Academy of Sciences. Of the group at breakfast that morning six, including Thomas, have gone; there are only three of us left: Crodowaldo Pavan, Hector Croxatto and myself.

Professor Odhiambo was awarded the Abdus Salam Medal of the Third World Academy of Sciences; and it was in July 1985 at the inauguration of TWAS in Trieste, where a number of distinguished African scientists were present, that he proposed the idea of an African Academy of Sciences, (of which he was Founder President), with the same objectives that drove him in his life, of putting science to work on real life problems. That was always the passion of his life. The most fitting tribute we can pay to our friend, a great scientist and servant of humanity, Thomas Odhiambo, would be to rededicate ourselves to the objectives, which have also been a major thrust of this Academy.

Mambillikalathil G.K. Menon

GEORGE PORTER († 31.VIII.02)

George Porter (hereafter GP) was an outstanding physical chemist and highly-successful populariser of science. He died in hospital on 31 August 2002 in his eighty-second year, in the cathedral city of Canterbury, in Southern England, not far from his home in the hamlet of Luddenham. In 1967, for his invention and application of the technique of flash photolysis starting twenty years earlier, he shared with Manfred Eigen and Ronald Norrish the Nobel Prize in Chemistry, awarded that year for 'studies of extremely fast chemical reactions, effected by disturbing the equilibrium by means of very short impulses of energy'. He became a member of the Pontifical Academy of Sciences in 1974.

GP was born on 20 December 1920 in Northern England, in the village of Stainforth in the West Riding of the County of Yorkshire, where he spent the first eighteen years of his life. Stainforth was then one of many farming and coal-mining communities near the town of Doncaster – Roman *Danum*. His father was a local builder, who in his spare time served as a lay Methodist preacher and school governor; one of his grandfathers was a coal miner. Most of his contemporaries at the local elementary school would have entered full-time employment on reaching the statutory leaving age, then fourteen. But GP was amongst the lucky few who at eleven years of age qualified on academic grounds to enter the secondary grammar school in the nearby mining village of Thorne, where chemistry became his best subject. The award of a scholarship in 1938 enabled him to enrol in a bachelor's degree course in chemistry at the University of Leeds.

We note here, in passing, that local education authorities in the West Riding did their best to encourage academically-gifted children of modest financial means to enter secondary grammar schools. They also offered scholarships to those who eventually qualified to enter university degree courses. In his autobiography *Home is where the wind blows*, Fred Hoyle, the eminent astronomer and cosmologist, who hailed from the Bradford area, entertainingly explains how in the mid-1930s, having taken his bachelor's degree at Cambridge, he persuaded the West Riding authorities to continue financing his academic studies, so that he could pursue postgraduate research under the nominal supervision of Paul Dirac. Other eminent scientists hailing from the West Riding include (at least) three Nobel laureates – the ionospheric physicist Edward Appleton of Bradford, the nuclear physicist John Cockcroft of Todmorden, and the inorganic chemist Geoffrey Wilkinson, also a native of Todmorden.

GP's interest in chemical kinetics was stimulated by his studies at Leeds, where he completed his degree in 1941. In parallel with his chemistry studies he undertook an obligatory extra course in radio physics. This introduced him to the electronics and pulse techniques that he was to apply so effectively when, some years later, he became engaged at Cambridge in research in physical chemistry. The radio-physics course was designed to meet growing demands for suitably trained personnel made during the early years of the Second World War by the British Armed Services, as they made operational use of the newly-developed techniques of radar. As a radar officer in the Royal Navy from 1941 to 1945, GP saw active service in the Atlantic and Mediterranean theatres of war.

Armed with a chemistry degree and his wartime experience of radar, in 1946 GP started graduate research towards a PhD under the supervision of Professor Ronald Norrish in the Department of Physical Chemistry of the University of Cambridge. Within a few years he had developed and started applying his highly original technique of flash photolysis, thereby revolutionising the study of fast chemical reactions. Throughout the whole of his subsequent academic career – from 1955 to 1966 at the University of Sheffield (also in the West Riding of Yorkshire) as Professor of Chemistry; from 1966 to 1985 at the Royal Institution in London as Director (in succession to Lawrence Bragg), Fullerian Professor of Chemistry and Director of the Davy Faraday Research Laboratory; and from 1985 until the end of his life at Imperial College London as Professor of Photochemistry and Chairman of the Centre for Photomolecular Sciences – GP was to lead or remain involved with research groups working in the area of physical chemistry that he had helped pioneer.

To paraphrase the account of GP's scientific work given in the Academy's Yearbook for 2001: – 'My research over the last forty years has been mainly in the area of photochemistry and the study of very fast chemical reactions. In 1949 I introduced the technique of flash photolysis for the study of chemical events in the microsecond region and, over the years, have extended this, by using pulsed lasers, into the nanosecond and picosecond regions. These methods led to observations of many new free radicals and the triplet states of both gases and liquids. I also introduced the technique of 'trapped atoms and radicals in a glass cage', which subsequently became known as 'matrix isolation'. Recently my principal application of these techniques has been to the primary processes of photosynthesis. Studies *in vivo* have elucidated the mechanism of light harvesting and I have devoted much attention to making models of the photosynthetic system *in vitro* with the ultimate objective of providing a practical artificial system for solar energy collection and storage'.

GP enjoyed the celebrity status resulting from his Nobel Prize award and his success as a popular lecturer, exploiting it in his dealings with politicians and government officials when expressing forthright views on science and education. Reputedly seeing the search for knowledge as the highest aim of mankind, he criticised proposals for concentrating research in selected centres on the grounds that this might 'stifle the original mind and encourage the safe and mediocre', and he attacked the spread of an anti-science lobby in the United Kingdom, condemning what he saw as 'skimping on long-term research and the concentration of scientific spending on short-term get-rich-quick projects'.

As Director of the Royal Institution – with its tradition going back to Humphry Davy and Michael Faraday of providing stimulating popular lectures by leading scientists – he was ideally placed for promoting public appreciation or understanding of science. A natural communicator and a pioneer of scientific programmes on television, he could engage the attention of audiences ranging from schoolchildren to advanced research workers.

Amongst the many honours showered on GP in recognition of his research and wider contributions to science was an invitation to succeed biophysicist Andrew Huxley as President of the Royal Society of London, in which capacity he served from 1985 to 1990. Whilst continuing to fight for more and better research in basic science and better science education, he supported and expanded the Royal Society's initiatives in the public understanding of science, and he became involved in issues of human rights affecting scientists in China, the Soviet Union and elsewhere. In 1990, the year in which he handed over the presidency of the Royal Society to pure mathematician Michael Atiyah, he was created Baron Porter of Luddenham and added service on the Select Committee on Science and Technology of the House of Lords to his many other activities.

I knew GP largely by reputation. Both of us hailed from mining communities near Doncaster and attended local elementary and grammar schools, but he was my senior by nearly a decade and our paths crossed no earlier than the mid-1970s. Our main dealings took place a decade later, in connection with Gresham College in the City of London. On my taking up the part-time professorship of astronomy there in 1985, I was asked to suggest measures for improving the College's effectiveness. To this end I started considering what practical steps would be needed to expand the scope of the College's activities through collaborative ventures with other bodies and to restore some of the College's traditional links with the Royal Society. Dedicated since its foundation in 1597 to the promotion of the appreciation of the arts and sciences, Gresham College continues to offer free public lectures in London, but it needs partners in carrying out its activities in the highly-specialised modern world, where the important subject of astronomy is just one of the physical sciences. I was grateful at the time for GP's helpful encouragement and advice in this matter. Associated with the resulting reforms was the appointment of a bemused GP several years later to serve (from 1990-93) as the thirtieth Gresham Professor of Astronomy.

When writing this commemorative address, I was conscious that my own field of science, geophysics, gives me no special qualifications for commenting usefully on George Porter's personal research work. This I leave to others such as Graham Fleming and David Phillips, who have prepared a detailed biographical memoir for publication next year by the Royal Society. In what I was able to write I relied heavily on commentaries by many of his friends and colleagues, to whom I must express my gratitude.

With the passing of George Porter, chemistry lost an outstanding practitioner, and science as a whole lost an energetic expositor and champion of its causes.

Raymond Hide

MARCEL ROCHE († 3.V.03)

Born in Caracas, Venezuela on August 15, 1920 – he died in Miami, Florida on May 3, 2003. He was Member of the Pontifical Academy of Sciences, appointed in 1968.

He had an international education being able to speak seven languages: Spanish, English and French fluently, since he learnt them at home and later he could also use properly Portuguese, Italian, Latin and Greek. He went to Senior High School in St Croix de Neville in Paris and graduated from it in 1938. He then studied Biology and Chemistry at St Joseph's College in Philadelphia, from which he graduated in 1942. He then entered the School of Medicine in Baltimore, Maryland, where he obtained a MD degree in 1946.

He spent a year as an intern in Johns Hopkins Hospital between 1946-47. He also worked as Residence Assistant in Medicine at Peter Bent Brigham Hospital in Boston from 1947 to 1948.

In 1953 he obtained a Medical Degree at the Central University of Venezuela and in 1970 spent a semester at Cambridge University. As a graduate student, he was an intern in the Department of History and Sociology of Sciences at the University of Sussex, England, from 1972 to 1973. After graduation he was Assistant Professor of Semiology between 1952 and 1956. In 1958 he obtained a Professorship in Physiopathology at the Central University of Venezuela where he remained until retirement. He held the 'Simon Bolivar for Latin-American Studies' Chair at the University of Cambridge 1970-71. He was also Research Fellow at the Department of Political Investigations of Sciences – University of Sussex 1973-74.

Administration

Professor Marcel Roche was the Founder and Director of the Institute for Medical Research in Venezuela (1952-58), Director of the Venezuelan Institute for Neurology and Brain Research (1958-59) and Director of the Venezuelan Institute for Scientific Research (1958-60). He was 'Gobernador' of the International Agency of Atomic Energy (1958-60), President of the Council of the United Nations University (UNU) 1978 and had 24 similar positions which he operated with success.

He had a list of over 30 awards comprising Premium Order, Gran Cruces, Doctor Honoris Causa. Besides the sciences, several of them were related to popularisation of Science, Sociology and Art.

Publications

In Endocrinology and Metabolism he has a list of 40 publications; in Endemic Goitre 19 publications; in Rural Anaemia 33 publications; in Ancylostoma Experimental studies 9 publications; in Humanities in Sciences 53 publications; 12 publications of Books and Monographs; Prefaces of 6 books.

Marcel Roche was a great medical scientist, a great intellectual citizen, and an excellent colleague. We deeply miss him.

Crodowaldo Pavan

SELF-PRESENTATIONS OF THE NEW MEMBERS

SUZANNE CORY

Being elected to the Pontifical Academy is a great and unexpected honour and I am awed to be here today in this historic place with such distinguished colleagues.

I was born and educated in Melbourne, Australia, and graduated in biochemistry from the University of Melbourne. During that time I became fascinated with DNA and the new science of molecular biology. With the naive audacity of youth, after I finished my Master's degree, I applied for PhD studies to the MRC Laboratory of Molecular Biology in Cambridge, where Watson and Crick had determined the structure of DNA. Crick accepted me into his department to work with Dr Brian Clark on the sequence of a transfer RNA, one of the adaptor molecules that decode the language of DNA into the language of proteins. This project enabled me to learn the powerful new sequencing techniques that had just been worked out by Fred Sanger.

My other great good fortune at Cambridge was to meet Jerry Adams, an American who had just completed his PhD with Jim Watson at Harvard, who was to become my husband and my lifelong scientific partner. After Cambridge, Jerry and I went to the University of Geneva, where we studied RNA from a small bacteriophage as a surrogate for messenger RNA, the molecular go-between that enables DNA recipes to be converted into proteins, the working molecules of our cells. Very surprisingly, we found that messenger RNA had a substantial amount of structure, due to the formation of base-paired loops, and also contained regions devoid of protein-coding potential.

In 1971, we returned to Melbourne to the Walter and Eliza Hall Institute, Australia's pre-eminent medical research institute and a world centre for immunology and hematology. Our first notable finding there was that mammalian messenger RNAs started with very bizarre structures, dubbed 'caps', which were later shown to act as landing pads for ribosomes to start protein synthesis. The development of recombinant DNA technology enabled us to clone messenger RNAs for antibodies and our laboratory helped to determine the number and arrangement of antibody gene elements in the germline and to show that lymphocytes 'cut and paste' these elements to create functional antibody genes.

Our interests then shifted to the molecular biology of cancer. We showed that the chromosome translocations hallmarking Burkitt's lymphoma activate the c-*myc* oncogene by fusing it to an antibody gene locus and then proved in transgenic mice that *myc* deregulation is profoundly leukemogenic. Following this paradigm, putative oncogenes have been identified from numerous translocations.

Today our laboratory is still preoccupied with the genetic accidents underlying cancer and our current major focus is to understand how cells decide whether to live or die. In 1988, David Vaux made the seminal finding in our laboratory that *bcl-2*, the gene activated by chromosome translocation in human follicular lymphoma, promotes cell survival. Together with Andreas Strasser and Alan Harris, we then found that co-expression of *bcl-2* and *myc* is profoundly lymphomagenic. These discoveries opened an entirely new way of thinking about cancer development – one of the critical steps is to block the normal process of cell death (apoptosis). Together with several other laboratories at the Hall Institute, we are dissecting how *bcl-2* and related genes regulate apoptosis and attempting to use this knowledge to develop more effective therapeutics for cancer and degenerative diseases.

ANTONIO GARCÍA BELLIDO

I am most honoured by the nomination of Member of the prestigious Pontificia Academia Scientiarum. I was born in Madrid (Spain) in 1936 where I studied in its Central University and obtained the PhD degree in 1962. My formative years include long stances in the Departments of Zoology, Cambridge, (UK), Zurich (Switzerland) and CalTech, California (USA). In 1969 I organized my own laboratory in the Spanish Research Council where I am now Professor of Research. My field of inquiry is Developmental Genetics and my model organism of research is *Drosophila melanogaster*.

My work has dealt with the general problems of how the genome, transmitted along different cell lineages during development, organizes the embryo, leads to distinct spatial patterns of cell differentiation and to species-specific morphogenesis. I introduced and elaborated the method of clonal analysis to analyze development at the cellular level. My research group studied the fundaments of somatic cell genetics for the experimental use of genetic mosaics Clonal analysis allowed us to describe development in terms of cell lineages what revealed the existence of 'developmental compartments', polyclonal modules of development, universal in animal morphogenesis. Genetic mosaics have permitted us to directly study mutant cell behaviours, caused by lethal mutations in the majority of the *Drosophila* genes. These cell behaviours reveal functions related to cell-cell recognition, cell induction, cell migration, patterned organization, controlled cell proliferation and hence size and shape of organs.

We analyzed in some detail over the years three model morphogenetic systems at the genetic, cellular and molecular levels. One is related to the genetic specification of embryonic territories, such as segments and compartments by the *Ubx* and *en* genes, which control the activity of other genes and were so called by us 'selector' genes. They act cell-autonomously defining specific cell behaviours of downstream 'realizator' genes. A second morphogenetic system has been the *ac-sc,* a gene complex involved in the specification of cell lineages leading to sensory neurons in invariant patterns. The third system deals with the specification of the lineal pattern of veins in the wing, tightly associated with its constant size and shape. The control of cell proliferation leading to these species specific parameters is still a major challenge in morphogenesis.

I have received honorary degrees by the Russian Academy of Sciences and various Spanish Universities. I am associated or foreign member of prestigious Academies like those of Sciences in Washington, in Paris and the Royal Society of London and ordinary member of the Spanish one. I have received Prizes like the L. Meyer of the Academy of Sciences of Paris and the Ramón y Cajal and Prince of Asturias of Spain.

WILLIAM D. PHILLIPS

Thank you, Mr President and members of the Academy. It is a great honour to be included in this membership. Today, November 5th, I am 56 years old, and this is a wonderful birthday present. I am looking forward very much to meeting more of you, to learning more about you and your work, and to benefiting from those experiences.

I was born 56 years ago. My parents were from very different backgrounds. My mother had emigrated from Italy to the United States when she was eight years old while my father's ancestors had come to the United States two centuries before. Both of my parents were the first people in their families ever to go to college or university and they transmitted to me and my sister and brother the great importance of education. In addition to that, they also imparted to us the importance of treating all people, everywhere, regardless of their backgrounds, with respect and with kindness. This stemmed not only from their own experiences but also from their deep religious convictions.

I was educated in public schools in Pennsylvania, that is, in state schools. My parents, although they were not scientists, encouraged my childhood interest in science, which turned towards physics by the time I was ten years old. Even though I did not really know what a physicist did, I was somehow convinced that I wanted to be one. I went to college at a small liberal arts college, Juniata College in Central Pennsylvania and then to graduate work at MIT. There I studied with Dan Kleppner, who really taught me how to think like a physicist. When I finished my degree at MIT, I spent two years there as a Chaim Weizmann Postdoctoral Fellow before I went to what was then the National Bureau of Standards, outside of Washington DC, now called the National Institute for Standards and Technology.

I had been hired to work at NBS/NIST because of my background in fundamental constants and precision measurement. At the beginning I worked on precise electrical measurements, but I was given the opportunity to spend a small fraction of my time playing around with some new ideas that were just becoming current at that time. Those ideas concerned how one could cool a gas of atoms using laser light. I remind you that in a gas the atoms and molecules are moving very rapidly if the gas is at ordinary temperatures. The nitrogen molecules in the air in this room are moving at about 300 metres per second. The idea that we were playing with at that time was to use laser light to push on the atoms in a gas so as to make them all slow down. Over a period of time, in my laboratory and in other laboratories, we learned how to accomplish this goal of slowing down the atoms in a gas so that the gas would become very cold.

At that time there was a theoretical understanding for how the laser cooling process worked, and that theoretical understanding led to a prediction of how cold the gas could be. In one of the most surprising and exciting experiments of my career, my colleagues and I discovered that, in fact, the temperatures we were reaching were much much colder than the theorists had told us were possible. There are not very many times in one's career that one can so definitively say that a very simple and elegant theory is completely wrong, but this was such a time. For me it illustrates that the criticism that is often levelled against science – that it is a closed system in which people simply reproduce the conventional ideas – is completely wrong. Any scientist would love to be able to disprove the most cherished ideas, because it is such an exciting thing to do. Well, that experiment led to a re-evaluation of the theory and our colleague Claude Cohen-Tannoudji, who will be speaking later about related ideas, and his colleagues at the Ecole Normale Supérieur in Paris, explained why it was that we were able to achieve such low temperatures.

With the insight gained from the new understanding about laser cooling we were able to go to even lower temperatures. In the end, we achieved temperatures less than one millionth of a degree above Absolute Zero. To emphasize how cold this is, consider that interstellar space is bathed in thermal radiation at about 3 degrees above Absolute Zero and we were achieving temperatures several million times lower than this coldest natural temperature in the universe.

What is all of this good for? Today, the unit of time is realized with atomic clocks that use laser cooled atoms. But the uses of laser cooling go beyond timekeeping. By putting these incredibly cold atoms into periodic structures formed by the interference of laser beams, we can create a kind of artificial laboratory for studying solid state physics. With these artificial solids we have been able to see phenomena that are not able to be seen in real solids. Another new area that we are exploring is the use of these individual ultracold atoms as bits for quantum information. Quantum information is a new kind of information science in which bits are stored and processed not as 'zeros' and 'ones', as they are in all of our laptop computers, but as a quantum mechanical superposition of 'zero' and 'one'. These quantum bits, or 'qubits', can be both zero and one at the same time. This opens up incredible and almost unimaginable new possibilities for processing information, ideas that are just beginning to be understood and are very far from being realised in the laboratory. Right now we are taking the first steps toward this new kind of information science.

As my career has developed, one of the things that I have been most mindful of is that anything good that I have ever done has been accomplished in the company of colleagues. Therefore, I am very happy to have an expanded set of colleagues by being part of this Academy of Sciences. I am very grateful for all the colleagues who have influenced me and from whom I have learned so much throughout my entire career. From the very beginning I benefited from my parents encouraging my early interest in science, and I have continued to be sustained by the love and support of my family. My wife, Jane, who is here today, has been extremely important in supporting my scientific endeavours and I am thankful to her and to all the members of my family. Finally, I am thankful to God for giving us such an incredibly beautiful and wonderful universe to explore and in which to have fun.

Yves Quéré

Mon père était originaire de Bretagne, ma mère de Lorraine (cas inverse de celui de Victor Hugo, mais là s'arrête la comparaison). C'est probablement de lui que j'ai hérité à la fois ma familiarité tactile avec la matière et ma fascination pour l'infinitude des horizons marins; d'elle l'irrépressible besoin d'une musique omniprésente comme celui des silences habités de la haute futée meusienne.

Mes études terminées, j'ai eu la chance de rencontrer un maître incomparable en la personne de Jacques Friedel. De lui m'est venu mon attrait pour la physique des solides et notamment celle des matériaux. J'ai travaillé – d'abord au Commissariat à l'Énergie atomique, puis à l'École polytechnique – sur les défauts cristallins et les effets des irradiations sur les solides, puis sur l'interaction des particules avec la matière et notamment sur les effets de la cristallinité dont la *canalisation* des particules constitue un exemple classique.

À l'École polytechnique, je suis devenu Président du Département de Physique puis du Sénat des Professeurs, enfin Directeur de l'Enseignement. Membre de l'Académie des Sciences depuis 1991, j'en ai été jusqu'à 2003 le Délégué aux Relations Internationales, ce qui me vaut aujourd'hui de coprésider l'*InterAcademy Panel* (IAP) qui regroupe l'ensemble des Académies des Sciences de par le monde (dont l'Académie Pontificale). Avec Georges Charpak et Pierre Léna, je participe depuis une dizaine d'années à cette rénovation de l'enseignement des sciences à l'école qu'est *La main à la pâte*, avec de multiples résonances dans nombre de pays étrangers.

Je ne saurais évoquer ce qu'a été ma carrière sans y associer la mémoire de France Quéré, mon épouse, décédée en 1995 à 58 ans. Écrivain, théologienne, auteur d'une œuvre considérable, conférencière invitée en mille lieux, Rome notamment, membre du Comité national d'éthique où son mandat fut – fait exceptionnel – renouvelé sans interruption, profondément habitée par les problèmes moraux que soulèvent la science en général et les neurosciences en particulier, elle a orienté en grande partie ma propre vie par son style, ses convictions, et ce mélange de rationalité et de lyrisme qui a marqué tant de ses lecteurs et de ses auditeurs.

VEERABHADRAN RAMANATHAN

My fundamental interest is in understanding how atmospheric gases, clouds and aerosols regulate the planetary greenhouse effect, solar radiative heating, climate and water budget. As a post doctoral fellow I identified the greenhouse effect of the vibration-rotation bands of chlorofluorocarbons (CFCs); on a per molecule basis, CFCs were about 10,000 times more effective than CO_2 as a greenhouse gas. This surprising finding opened the door to the discovery of the greenhouse effect of numerous other trace gases and the field of trace gases-climate-chemistry interactions. Subsequently, I identified the global cooling effect of stratospheric ozone reduction and that anthropogenic increase in tropospheric ozone can have a strong greenhouse effect. Showed that non-CO_2 trace gases, including, CFCs, Methane and Nitrous oxide, are just as important as CO_2 for global warming. The trace gas work culminated in a WMO sponsored assessment report (chaired by me) on Trace Gases Effect on Climate which established the fundamental importance of non-CO_2 trace gases to the Global warming problem. Predicted in early 1980s that the global warming signal should rise above natural fluctuations by 2000.

Clouds, the Gordian knot of the climate problem, were my next focus. Using radiation budget measurements on board a NASA satellite, showed that clouds have a large natural global cooling effect on the planet, i.e, the reflection of solar radiation back to space exceeded the greenhouse effect of clouds. Showed that the atmospheric greenhouse effect can be obtained from satellite-measured outgoing infrared radiation data. Proposed that tropical high-thick clouds act as a thermostat for regulating surface temperatures of the warmest oceans in the planet and conducted the Central Equatorial Pacific Experiment from Fiji with aircraft, ships and satellites for testing the Thermostat hypothesis. The new finding of the experiment was that the cloudy atmosphere was absorbing significantly more solar radiation than predicted by theory and models. Suspecting nano-meter to micrometer size atmospheric particles (aerosols) may provide the clue to this absorption, I designed (with P.J. Crutzen) the Indian Ocean Experiment, and discovered (in collaboration with 200 scientists) the S. Asian brown cloud and its significant environmental impacts. Also provided the first observational demonstration that the highly absorbing S. Asian haze significantly reduces sunlight reaching the Earth's surface over a very broad region. The INDOEX results linked human activities in the heavily populated S. Asian area to key aspects of regional climate including rainfall.

This work led to an UNEP initiated project, Atmospheric Brown Clouds (ABC), to study the impact of such brown clouds worldwide.

At the Scripps Institution of Oceanography of the University of California, San Diego, I am a distinguished Professor of atmospheric and climate sciences and the Director of the Center for Atmospheric Sciences. I am also the co-chief scientist of the Atmospheric Brown Cloud Project and the Indian Ocean Experiment.

THE PIUS XI GOLD MEDAL AWARD

LAURE SAINT-RAYMOND

Summary of Scientific Activity

All along my degree course, I could not bring myself to choose between mathematics and physics, torn between the desire to understand the world that surrounds us and the wish to build some more abstract framework: in the Ecole Normale Supérieure, I could notably enjoy a double training. As a daughter of mathematicians, I have finally opted for the rigour of mathematics but without losing sight of applications. Actually my research is more precisely motivated by plasma physics and meteorology.

My first steps as a researcher, under the supervision of François Golse, were turned onto the study of charged particles submitted to strong constant external magnetic fields (tokamaks, plasmas constituting planetary environments...). From purely mathematical point of view, these works have introduced me to two theories which are still in the heart of my scientific activity: the kinetic theory which allows to model rarefied flows, and the problems of singular perturbations which consist to derive simplified models from a given system in some particular asymptotic regimes. From the point of view of applications, these works have allowed to give a rigorous multiscale analysis of the motion of such plasmas, which is a superposition of fast magnetic oscillations, of an electric transport in the direction of the magnetic field and of a slow drift due to the coupling between magnetic oscillators.

All these results can be easily transposed to the problem of rotating fluids, which is to the study of fluids submitted not to magnetic fields but to the Coriolis force, which is a problem of first order in meteorology. Nevertheless, at some latitudes, the curvature of the Earth has to be taken into account and the approximation of constant penalization for the Coriolis force becomes very false. A series of joint works with Isabelle Gallagher aims at understanding the influence of these homogeneities and in particular to rediscover some trapping phenomena occuring in the equatorial zone.

Among these works, this is probably the one which is the most distant from physicians' concerns which has been accorded most attention ... of mathematicians. It is indeed a partial answer, obtained in collaboration with François Golse, to a question asked by Hilbert on the occasion of the International Congress of Mathematicians in 1900. The problem consists actually in getting a unified description of gases from the molecular scale to the macroscopic scale. The mathematical treatment of such a problem can seem if not out of subject, at least unnecessarily complicated, but it allows actually to find in a systematic way some well-known physical phenomena.

From a physical point of view, the problem of fluid limits consists in understanding the links between the various levels of modelisation. The classical models introduced by Euler for inviscid fluids (18th century) and by Navier and Stokes for viscous fluids (19th century) describe fluids as 'continuous' media, thanks to a small number of measurable thermodynamic variables (temperature, pressure, bulk velocity…). The evolution of the fluid is then governed by a system of partial differential equations expressing the local conservations of mass, momentum and energy, system which is closed by a state relation. Such an approach consists then in considering infinitesimal volumes of fluid and to write balance equations for these fictitious particles. This phenomenological approach is actually compatible with the microscopic dynamics of molecules in the fast relaxation limit.

Indeed at the microscopic level the fluid is constituted of a large number of interacting particles governed by Newton's principle. Statistically the collisions between these particles induce a relaxation mechanism: if these collisions are sufficiently frequent, local thermodynamic equilibrium is reached almost instantaneously, which imposes a relation between the various thermodynamic variables, the so-called state relation. And the macroscopic equations of motion are nothing else than the averaged microscopic equations. In the particular case of perfect gases, the volume occupied by the molecules is negligible compared with the volume of the fluid: only binary collisions between uncorrelated particles have a determining role in the evolution of the fluid. In other words a perfect gas can be described by a statistical approach, thanks to its distribution function which gives the instantaneous number of particles of any fixed position and velocity. This distribution function is then governed by a partial differential equation of Boltzmann type.

In the case of perfect gases, the problem of fluid limits asked by Hilbert can therefore be decomposed in two subquestions, the derivation of the Boltzmann equation from Newton's principle (problem solved by

Lanford in 1974 for small times) and the derivation of hydrodynamic limits of the Boltzmann equation which is the matter of the works that I would like to present here.

The formal study of hydrodynamic limits of the Boltzmann equation goes back to Hilbert for inviscid perfect gases, and to Chapman and Enskog for slightly viscous perfect gases (the viscosity of perfect gases being necessarily small since the size of particles is negligible). This study lies on the fundamental features of the Boltzmann equation which takes into account both the transport of particles and the statistical effect of instantaneous elastic collisions. The symmetry properties of the collision operator imply actually the local conservations of mass, momentum and energy, and the local increase of some quantity, the so-called entropy. For fixed mass, momentum and energy, the entropy is maximal (and the collision operator cancels) when the velocities of particles are distributed according to a Gaussian, as predicted by Boltzmann.

This means in particular that, if the collisions are sufficiently frequent, the entropy increases rapidly and the distribution of velocities relaxes rapidly to a Gaussian. The state of the gas is therefore completely determined by its thermodynamic fields, which are the temperature, the macroscopic density and the mean velocity. The hydrodynamic equations are then obtained as approximations of some averages of the Boltzmann equation in the fast relaxation limit, that is when the Knudsen number (measuring the ratio between the mean free-path and the typical observation length) is very small.

Depending on the relative sizes of the Knudsen number and of the Mach number (measuring the ratio between the bulk velocity and the thermal velocity), the evolution of the gas is described by different hydrodynamic equations. The flow is compressible if the Mach number is of order 1, and incompressible if the Mach number is small. In this last case, the flow is inviscid if the Mach number is large compared with the Knudsen number, and viscous if both parameters are of the same order of magnitude: indeed for perfect gases the Reynolds number (measuring the inverse viscosity) is linked to the Mach and Knudsen numbers through the Von Karman relation. The mathematical study of each one of these asymptotics is very similar to the study of the corresponding hydrodynamic model: the hydrodynamic approximation leading to the incompressible Navier-Stokes equations is therefore less difficult to understand.

The main diffficulties encoutered to make rigorous the formal derivation sketched previously are actually linked to the physics of the system. The first problem consists in getting a control on particles of high energy: one has to check that such particles are not 'lost asymptotically', or in other words that there is no loss of energy transported by particles which would escape almost instantaneously from the observation domain. In mathematical terms, such a phenomenon would be translated into a loss of compactness (with respect to velocities) on the distribution function, or more precisely on its second moment. The usual a priori estimates based on the entropy and energy bounds do not allow to exclude such a scenario: refined a priori estimates based on the entropy dissipation bound are required to prove that, in the fast relaxation limit, the distribution function has almost the same regularity with respect to velocities as the corresponding Gaussian.

The passage to the limit in the moment equations associated to the Boltzmann equation requires furthermore a precise study of the oscillations created in the system. First of all, one has to check that there is no turbulent effect at the microscopic scale, that is no oscillatory behaviour on lengths of the order of the mean free-path which would destabilize the whole fluid by some resonance effect. The absence of spatial oscillations is obtained at the mathematical level by some compactness (with respect to positions) on the distribution function: averaging lemmas show indeed that the control on the advection (which is exactly balanced by collisions) gives some regularity (with respect to the space variable) on the thermodynamic fields which are nothing else than averages of the distribution function.

In the same way the system could be destabilized by temporal oscillations (on times of the order of the inverse sound speed, which is supposed to be small). Such oscillations actually take place in the fluid, they are known as acoustic waves (taking into account the weak compressibility of the fluid in such a regime). One has then to check that they do not produce constructive interferences, that is to describe precisely the propagation of the waves and their coupling in the moment equations. An argument of compensated compactness, due to Lions and Masmoudi allows then to conclude that the acoustic waves do not modify the average flow, and consequently that they do not occur in the limiting hydodynamic system.

Besides a mathematical challenge, the rigorous study of hydrodynamic limits is then a way to understand sharp physical phenomena taking place in quasi-hydrodynamic regimes. Therefore it finds out direct applications in the development of technologies requiring precise multiscale numerical simulations (aeronautics for instance). Of course, a great number of applications – like medicine – would require a more extended study, considering in particular complex microscopic interactions between elementary particles (blood or breath flows).

HOMILIE AM 32. SONNTAG IM JAHRESKREIS IN MONTECASSINO (7. 11. 2004) FÜR DIE PÄPSTLICHE AKADEMIE DER WISSENSCHAFTEN

JOSEPH CARD. RATZINGER

2 Makk 7,1-2.7a.9-14;
2 Thess 2,16-3,5;
Lk 20, 37-38.

Im November, während wir das große Sterben in der Natur erleben, spricht die Liturgie von unserem eigenen Sterben. Zur Frage nach dem rechten Leben gehört auch die Frage nach dem rechten Sterben. Wenn wir den Tod verdrängen müssen, dann werden wir auch mit dem Leben nicht recht umzugehen lernen. Leben und Sterben gehören zusammen, das Leben kann nur gelingen, wenn wir auf rechte Weise dem Tod entgegenzugehen vermögen.

Was ist es mit dem Tod? Was bedeutet er über seine biologische Gesetzlichkeit hinaus für die Ganzheit unseres menschlichen Lebens? Im Alten Testament hat sich eine Antwort darauf erst langsam entwickelt – Gott führt den Menschen sozusagen behutsam, Schritt für Schritt in die Geheimnisse des Lebens ein, zu denen das Geheimnis des Todes gehört. Die Vorstellung in den frühen Büchern des Alten Testaments ist derjenigen sehr ähnlich, die wir beispielsweise bei Homer für die griechische Welt vorfinden können. Im Tod steigt danach der Mensch ins Reich der Schatten hinab – irgendwie gibt es noch etwas von ihm, aber diese Existenz ist eine Un-Existenz, mehr Nichtsein als Sein. Die eigentliche Antwort, die Gott den Menschen geben wollte, ist nur langsam in Sicht gekommen: Sie ist im Ringen mit dem Leid im Beten Israels langsam gereift. Sie hat zunächst noch gar keine feste Gestalt, keine Form einer philosophischen Anthropologie, sondern ist nur als unfaßbare und gerade so erschütternde und heilende Gewißheit im Stehen vor Gott, im Sprechen mit ihm inmitten einer unverständlichen Welt aufgeblitzt. Ich bringe dafür zwei Beispiele.

Das eine ist der Psalm 73, ein Lieblingspsalm des heiligen Augustinus, der darin sein ganzes eigenes Fragen, Leiden und Hoffen wiedererkannte. Die frühe Weisheit Israels war davon ausgegangen, daß es dem guten Menschen – wenn auch durch Phasen der Prüfung hindurch – gut geht, und daß die Schlechtigkeit den Menschen ruiniert, ihre Strafe sozusagen in sich selber trägt. In den Schrecknissen der Geschichte Israels als Volk wie im Leiden der einzelnen gläubigen Menschen war dieser Optimismus allmählich unwiderruflich zerbrochen. Nein, die Hochmütigen, die Habsüchtigen, die Verächter Gottes sind die Erfolgsmenschen, sie sind reich und fett und können über den Gläubigen herfallen und ihn verhöhnen. Und die Gläubigen, die Gottes Willen folgen, nicht von der Wahrheit und nicht von der Gerechtigkeit abweichen – sie sind die Marginalisierten in der Geschichte, deren Leben Jesus im Bild des armen Lazarus zusammengefaßt hat, der vor der Tür des Reichen sitzt und schon dankbar wäre für die Brosamen, die vom Tisch des Reichen abfallen. Diese Erfahrung beschreibt der Beter des Psalmes 73 – sie ist seine Lebenserfahrung. Am Ende fragt er: Habe ich also umsonst mein Herz lauter bewahrt? (Vers 13). Er geht in den Tempel, um zu beten, und nun wird ihm Einsicht: „Als ich mein Herz verbitterte... da war ich töricht und ohne Einsicht, wie ein Riesenvieh vor dir. Nun aber bin ich immer bei dir... Wen habe ich im Himmel neben dir? Neben dir begehre ich nichts auf Erden. Mein Herz und mein Fleisch schwinden dahin, aber mein Anteil bleibt Gott auf ewig“ (Vers 21-26). Der Beter macht sozusagen die Erfahrung der Absolutheit der Liebe: Das Gute über allen Gütern ist das Geliebtsein von Gott, das nicht vergeht. Es ist das eigentliche Gut. Die anderen Güter kommen und vergehen, sie erscheinen nun in ihrer ganzen Relativität. Das wirkliche Gut ist es, mit ihm zu sein, von seiner Hand gehalten Und diese Hand läßt mich nicht los. Da ist kein Neid mehr nötig auf das Glück der Reichen. „Die Nähe zum Herrn ist mein köstliches Gut“ (Vers 28) – kein Lohngedanke ist da, der das Gute nur tun will, weil es sich lohnt, sondern einfach das Frohwerden im Angesicht dessen, der als das wirkliche Gut erfahren wird, das zugleich als unzerstörbar erkannt wird: Die Hand Gottes hält mich auf immer, im Tod und im Leben.

Die zweite Stelle, die ich erwähnen möchte, ist das Hoffnungsbekenntnis des Ijob in einem Abgrund von Leiden. „Erbarmt, erbarmt euch, ihr meine Freunde, denn Gottes Hand ist’s, die mich traf. Warum verfolgt ihr mich wie Gott, seid unersättlich ihr nach meinem Fleisch?“ (19,21f). In dieser Situation, in der Ijob von allen verlassen und verachtet sein Leben verwünscht, bricht sein Glaube an den wirklichen, den verborgenen Gott durch: Er appelliert an den wirklichen Gott gegen den verfolgenden Gott, und da wird ihm eine wundervolle Gewißheit geschenkt: „Ich weiß, daß mein Erlöser lebt... Ohne meine Haut, die so zerfetzte, und ohne mein Fleisch werde ich Gott schauen... Meine Augen werden ihn sehen..." (19,25ff). In der Hölle seines Leidens wird es dem betenden und glaubenden Ijob gegen allen Anschein, gegen den Gottesschrecken, der ihn befallen hat, klar: Ich weiß, daß mein Erlöser lebt, und ich werde ihn schauen. Ich weiß, daß der Gott, der mich scheinbar quält, in Wahrheit der erlösende Gott ist, auf den ich setzen darf und dessen Liebe mich durch die Nacht des Leidens und des Todes hindurchträgt.

Ich glaube, es ist wichtig zu sehen, daß im Alten Testament nicht zunächst aus einer durchgeführten Anthropologie heraus ein Unsterblichkeitsglaube entsteht, sondern daß die Begegnung mit Gott, dem unbegreiflichen und doch zutiefst verläßlich guten Gott, dem Menschen den Halt gibt, der ihn auch über den Tod hinüberträgt und der ihm daher auch im Leben den rechten Weg zeigt.

Erst in den späten Schichten des Alten Testaments, bei Daniel und in der Jesaja-Apokalypse erscheint dann ganz klar die Hoffnung auf die Auferstehung, die freilich im einzelnen weder ihrer Ausdehnung noch ihrer Art nach beschrieben wird. Die Kraft dieses Auferstehungsglaubens sehen wir in der Lesung aus dem späten zweiten Makkabäer-Buch, die wir eben gehört haben: Die Gewissheit der Auferstehung wird die Kraft zum Widerstand gegen den Tyrannen, sie wird die Kraft zum guten Leben und die Kraft, auch um den Preis des eigenen Lebens zum Wort Gottes zu stehen, weil dieses Wort eben die eigentliche Macht ist, die Leben schenkt, das wirkliche Leben über den Tod hinaus und jenseits des Todes.

Das Ringen um die Frage nach Tod und Leben ging in Israel freilich weiter – es ist ja im Letzten auch unser nie ganz abgeschlossenes Ringen. Auch wir müssen immer neu und in immer neuen Lebenszusammenhängen die Antwort erlernen, so daß sie unser Leben formen kann. Das Evangelium dieses Tages läßt uns einen wichtigen Ausschnitt dieses Ringens sehen und schenkt uns die Antwort Jesu, deren Tiefe wir freilich auch erst immer neu ertasten müssen. Das Evangelium zeigt uns die zwei Hauptpositionen des Judentums der Zeit Jesu. Da ist zum einen die Priesteraristokratie der Sadduzäer, die zugleich traditionalistisch und rationalistisch denkt. Sie sieht nur die Thora, die fünf Bücher Mose als kanonisch an und lehnt daher die späteren Entwicklungen der Glaubensgeschichte Israels ab, zu denen auch der Auferstehungsglaube gehört. Dieser war hingegen bei den Pharisäern und auch in breiten Volksschichten bestimmend und hat vor allem im Volksglauben ganz ähnlich wie später im islamischen Volksglauben phantastische und grob sinnliche Züge angenommen. Ein gefeierter Schriftgelehrter meinte beispielsweise: Dereinst – nach der Auferstehung – wird die Frau jeden Tag gebären. Die Auferstehungswelt erscheint so als eine geradezu ins Sinnlose übersteigerte Verdoppelung dieser Welt. Dagegen konnten die Pharisäer leicht polemisieren. Das Evangelium zeigt uns ein Beispiel, wie sie einen so entstellten Auferstehungsglauben ins Lächerliche zogen. Weil sie nur die fünf Bücher Mose als kanonisch anerkannten, mußte Jesus aus diesen Büchern argumentieren, um den Auferstehungsglauben zu rechtfertigen – was vom Textbefund her ganz aussichtslos erscheinen muß. Aber zunächst rückt Jesus die Vorstellungen von der Auferstehung zurecht. Die Auferstehungswelt ist nicht ein Doppel der Unseren: Fortpflanzung und Tod gehören zusammen; wo es keinen Tod gibt, gibt es auch keine Fortpflanzung mehr. Die Auferstandenen sind neue Menschen, „Gottessöhne" geworden. Sie leben in der Weise Gottes, im Sein vor Gott, mit Gott und zu Gott hin. Das „Sein wie Gott", das der Mensch im Paradies suchte und das er immerfort sucht – der Schrei nach völliger Freiheit in unserer Zeit ist ein Schrei nach Göttlichkeit – das ist ihnen gegeben. Ein solches Leben entzieht sich unserer Vorstellung, aber das Eine wissen wir, daß Gottes Sein wesentlich Wahrheit und Liebe ist. Dann ahnen wir auch, daß das künftige Leben einfach Angehaltensein an Wahrheit und Liebe und so Angehaltensein an Gott ist.

Gerade dies aber verdeutlicht der Herr in seinem Schriftbeweis für die Auferstehung. Mose nennt „den Herrn den Gott Abrahams, den Gott Israels und den Gott Jakobs... Er ist doch kein Gott von Toten, sondern von Lebenden, denn für ihn sind alle lebendig" (Lk 20,38). Diese Begründung der Auferstehung, des ewigen Lebens, ist überraschend. Der Herr begründet es von der Gottesgemeinschaft des Menschen und führt damit genau die Linie fort, die wir besonders im Psalm 73 gefunden hatten. Abraham, Isaak und Jakob haben als Freunde Gottes gelebt, im ständigen Gespräch mit ihm, im Mitgehen mit ihm, und so sind sie geradezu zu Namen Gottes geworden: Der Verweis auf sie klärt, um welchen Gott es sich handelt, wer Gott ist und wie Gott ist. Sie gehören zu ihm, und wenn sie Gott zugehören, wenn ihre Verbindung mit ihm das Wesentliche ihres Lebens ist, dann gehören sie dem Leben selbst zu. Weil sie an Gott angehalten sind, können sie nicht ins Nichts fallen. Sie leben ein Leben, das stärker ist als der Tod. Jesus gibt uns eine dialogische, eine relationale Begründung der Unsterblichkeit. Das Leben des Menschen reicht nicht deswegen über den Tod hinaus, weil etwa die Seele in sich unteilbar und darum unzerstörbar wäre, wie die griechische Philosophie die Unsterblichkeit begründete. Das In-Beziehung-Stehen macht den Menschen unsterblich. Menschliche Liebe ist auf Unendlichkeit angelegt, kann sie aber nicht geben. Die Liebesgemeinschaft mit Gott gibt, was der Liebe wesentlich ist: Dieser Dialog bricht nicht ab. Durch das Mitsein mit ihm sind wir im Eigentlichen, im unzerstörbaren Leben. Indem Jesus auf Abraham, Isaak und Jakob als Menschen verweist, die Gottes sind und daher lebendig sind, sagt er uns: Halte dich im Leben an das, was nicht vergeht und verfällt. Halte dich an die Wahrheit, an das Gute, halte dich an die Liebe, halte dich an Gott. Und von Christus selber her könnten wir nun sagen: Halte dich an den auferstandenen Christus an, dann hängst du am Leben, und dann lebst du wirklich – dann lebst du jetzt schon das wahre Leben, das ewige Leben.

Jesus lehrt uns also nicht irgendwelche mysteriöse Dinge über das Jenseits. Er weist uns ins rechte Leben ein. Seine dialogische Begründung der Unsterblichkeit sagt uns, wie wir jetzt leben müssen, wie wir das wahre Leben finden, das ewig ist. Was er uns von der Auferstehung sagt, ist durchaus praktisch; indem er den Tod deutet, zeigt er, wie das Leben geht. Von da aus können wir auch das Buch der Makkabäer neu lesen. Wer die rechten Güter, das wahre Gut – Gott – kennengelernt hat, der kann die relativen Güter fallen lassen. Der weiß, daß er auch das biologische Leben riskieren kann und darf, daß er nicht ins Nichts fällt, weil er so gerade das rechte Leben ergreift. Er weiß, daß das wahrhaft Gute der wahrhaft Gute ist und daß Er unsere Hand nicht losläßt. Bitten wir den Herrn, daß wir recht zu leben lernen. Amen.

HOMILY PREACHED BY CARDINAL JOSEPH RATZINGER AT MONTECASSINO ON 7th NOVEMBER 2004

2 Macc 7,1-2.7a. 9-14;
2 Thess 2,16-3,5;
Lk 20,37-38.

In November, as we experience dying in nature, the liturgy draws attention to our own death. To answer the question how to live well we also have to answer the question how to die well. If we suppress all thoughts of death, then we will never learn how to deal with life properly. Life and death belong together, life can only turn out well if we are able to approach death in a proper way.

What about death? What does death – which we all must suffer – mean for our human life as a whole? The answer to this question only gradually developed in the Old Testament: God gently introduced man into the mysteries of life, part of which is the mystery of death. What people thought about this subject in the early books of the Old Testament is very similar to what we can find in Homer, a representative of the Greek culture. According to him, in death man descends into the realm of shadows. There is still something left of man, but this existence is a non-existence, more 'not – existing' than existing. The real answer which God wanted to give man only gradually came to light. It matured in the prayers of Israel reflecting man's struggle with suffering. At the beginning there was no definite answer, no philosophical anthropology. But an answer appeared in his prayers in a shattering but also healing certainty amidst an incomprehensible world. I will give you two examples.

The first one is psalm 73, one of St. Augustine's favourite psalms, in which he recognised his own questions, sufferings and hopes. The early wisdom of Israel assumed that things would go well for the good man, despite periods of trial. Evil conduct would ruin him, because evil itself had punishment as a consequence. But the horrors which Israel had to bear in its history as a nation and in the sufferings of individual believers gradually destroyed this optimism irrevocably. No, the arrogant, the avaricious, people with contempt for God – these are the successful ones, these people are rich and fat, and they can attack and ridicule the believers. But the faithful who follow God's will and who do not deviate from truth and justice – these are the ones who are the insignificant people in history. Jesus summarised their lives in the example of Lazarus who lay at the rich man's gate and who longed to fill himself with the scraps that fell from the rich man's table. The person praying psalm 73 describes this experience – it is his personal experience of life. At the end he asks why he should have kept his own heart pure (verse 13). He goes to the temple to pray and so he realises: 'My soul was embittered… I was stupid and ignorant, like a brute beast in your presence. Nevertheless, I am continually with you… Whom have I in heaven but you? And there is nothing on earth that I desire other than you. My heart and my flesh may fail, but God is the strength of my heart and my portion forever' (verses 21-26). The one praying this psalm has the experience of being loved absolutely: The good above all goods is being loved by God, a love that is everlasting. This love is real wealth. All the other worldly goods come and go, and can be recognised as utterly relative. The real wealth is to be with Him, to be held in His hand. And this hand always holds me. I do not have to be envious of the good fortune of rich people. 'I have made the Lord God my refuge' (verse 28). The point is not that I will only do good if I am rewarded, but that the believer is joyful because he has experienced God as his true wealth, a wealth which cannot be destroyed: The Lord always holds him in His hand – in life and in death.

The second example I want to mention is Job's profession of faith in the midst of an abyss of suffering. 'Have pity on me, have pity on me, O you my friends, for the hand of God has touched me. Why do you, like God, pursue me, never satisfied by my flesh?' (*Job* 19,21-22). Although, abandoned and despised by everyone, Job curses his life; still his faith in the real, the hidden God, breaks through: he appeals to the real God against 'the pursuing God' and a wonderful certainty is given to him: 'For I know that my Redeemer lives… Then in my flesh I shall see God, whom I shall see on my side' (*Job* 19,25-27). In the hell of his suffering Job, the praying and faithful believer, realises clearly, in contrast to every impression given and in contrast to his 'horror of God' which has overcome him, that his 'Defender', his Saviour, lives, and that he will gaze on God. I know that the God who seemingly tortures me is in reality the Redeemer God whom I can trust and whose love carries me through the night of suffering and of death.

It is important to see that the belief in immortality, in the Old Testament, did not result from a fully worked-out anthropology. Rather, the encounter with God – the God who is incomprehensible but who can be trusted nevertheless – gave man the support he needed to carry him through death, and in this way pointed him to the right way in life.

Only in the later parts of the Old Testament, in Daniel and the Isaiah apocalypse, does the hope of resurrection clearly appear, though this hope of resurrection is not described in detail. We can see the power of this belief in resurrection in today's reading from the second book of Maccabees. The certainty of resurrection gives the power to resist the tyrant. It becomes the power to lead a good life and to stand by the Word of God even in the face of death, because this Word is the actual power which gives life, the real life over death and beyond death.

The struggle about the question of life and death continued in Israel. Indeed, it is actually our own struggle too, one that we can never fully complete. We, too, have to learn the answer anew in the ever new conditions of life so that this answer can mould our own lives. Today's Gospel makes us see an important part of this struggle and gives us our Lord's answer whose full meaning we always have to work out anew.

The gospel presents us with the two main positions of the Jews at the time of Jesus. On the one hand, there is the priestly aristocracy of the Sadducees who were in the same time traditionalists and rationalists. They accept only the five books of Moses as canonical and therefore reject the later developments in the history of faith in Israel, including belief in the resurrection of the dead. This belief, on the other hand, was crucial for the Pharisees as well as also for large segments of society. In popular belief, however, it took on fantastic and coarsely sensuous characteristics, very similar to Islamic popular belief. A celebrated scribe, for example, thought that in the future – after the resurrection – women will give birth to children every day. The world of resurrection thus looks like a senselessly excessive exaggeration of the present world. It was easy for the Sadducees to inveigh against these ideas. The gospel shows us an example how they ridiculed such a distorted belief in resurrection. As the Sadducees only accepted the five books of Moses as canonical, Jesus would have to rely on texts from these books for arguments to justify the belief in resurrection – arguments the texts themselves could not provide without difficulty.

First Jesus corrects faulty ideas about the resurrection. The world of the resurrection is not simply a exaggeration of our world: reproduction and death belong together; where there is no death, there is no longer reproduction. The risen people have become new beings, "Sons of God". They live in the manner of God, exist before God, with God and towards God. The "being like God" which man looked for in paradise and which he constantly looks for – the cry for absolute freedom in our time is a cry for being godlike – this desire is inherent in man.

Although such a life is beyond our imagination, there is one thing we know, namely that God's existence is essentially truth and love. We can also assume that the future life means to be in touch with truth and love and thus to be in touch with God. This is just what Jesus wants to explain when he uses the Scriptures to prove resurrection. Moses calls 'the Lord the God of Abraham, the God of Isaac and the God of Jacob … he is God, not of the dead, but of the living; for to him all men are in fact alive' (*Lk* 20,30). To give such a reason for resurrection, for eternal life, is surprising. The Lord finds the reasons for resurrection in the community of man with God and his reasoning is in line with that discerned in psalm 73. Abraham, Isaac and Jacob lived as friends of God, in constant conversation with Him, in going with Him. Indeed, they have actually become central to the way in which God is named: reference to them settles the question what kind of God we are talking about, who God is and what God is like. They belong to Him and if they belong to God, if their contact with God is the most essential part of their lives, then they belong to life itself. As they are in touch with God they cannot fall into void. The lives they live are more powerful than death.

In this way, Jesus supplies interpersonal and relational grounds for immortality. The reason why human life stretches beyond death is not because the soul is in itself indivisible and thus indestructible as Greek philosophy argues. It is the 'being-in-relationship' that makes man immortal. Human love wants infinity but cannot give it. The communion of love with God is essential to love: this relationship never breaks up. If we are with Him then we are in the real, the indestructible life. By referring to Abraham, Isaac and Jacob as human beings who belong to God and are thus alive Jesus tells us: In life hold on to the 'things' which do not perish and disintegrate. Hold on to truth, to the Good, hold on to love. Hold on to God. And by looking upon Christ we could now say: Hold on to the Risen Christ, then you hold on to life and then you live truly – because then you already live the true life, the eternal life now.

Thus Jesus does not teach us any mysterious things about the next world, but he guides us how to lead a true life. By adducing dialogic reasons to argue about immortality, Jesus instructs us how we have to live now, how we can find the true life which is eternal. What he tells us about resurrection is quite practical: by interpreting death he shows us how life has to be lived. Given this new perspective we can read the book of Maccabees in a new way. Whoever has got to know the really good things, the true Good – namely God –, can give up the relatively good things. Such a person knows that he can and may risk his biological life, that he does not fall into void, because by taking this risk he takes hold of the true life. He knows that the truly Good is God alone and He does not let go our hands. Let us pray to God that he may teach us how to live truly. Amen.

SCIENTIFIC PAPERS

DIFFERENT TYPES OF DISCOVERY LESSONS FROM THE HISTORY OF SCIENCE

JÜRGEN MITTELSTRASS

What is known to science is known in different ways – a quick look at, for instance, textbooks in mathematics, physics, biology or economics or history renders this apparent. The same is true of the ways in which science acquires its knowledge. There are inductive, deductive, experimental and many other ways and methods. And its successes are documented in confirmed hypotheses, explanations of the hitherto inexplicable, and in discoveries.

It is above all the discoveries which represent the appeal of science – for the scientific layman as well as for the scientist. The new is revealed in discoveries, and the new is, by all means, the aim of scientific endeavours. Aiming at the discovery of the new, science develops its methods and structures and defines its concept of research. Thus, occasionally a distinction is made between hypothesis-driven and discovery-driven research, apparently thinking of the expected new in the former, and of the unexpected new in the latter case. But such a distinction is artificial. After all, hypothesis-driven research is also aimed at discovering the new, and discovery-driven research also requires methodic tools, parts of which are in turn hypotheses. Precisely this is what the history of science teaches, provided one conceives it not just as an arsenal of past insights and errors, but also as an expression of the scientific spirit, which recognises itself in past as well as present representations.

Not even science has a subscription to the new or a strategy for finding it routinely. There are many roads not just leading to Rome, but to new scientific insights as well – and, of course, many leading past them too. I shall describe in more detail three paths through which new insights have been found in the history of science, using short examples: (1) Discoveries which were surprising to science and the discoverers themselves, (2) discoveries which had been expected by the discoverer and science up to a point, but which were novel in their details, (3) discoveries which had been expected by the discoverer, but which came as a complete surprise to science.[1]

First, the discoveries which came as surprises to science and even to the discoverer himself, hence representing something unexpected and new. Two examples: (1) The first is very famous: the *discovery of the X-ray*. In November 1895, Röntgen was experimenting with a gas discharge tube, that is, an instrument that emits electrons (for instance, from a heated wire), and accelerates them by applying voltage; the tube itself is under vacuum, or at least, the gas pressure is significantly reduced. He covered the tube with cardboard paper to see to what extent it would still let light pass, and thus observed that some crystals, left lying on his desk for no particular reason, started to fluoresce. Röntgen discovered that this fluorescence must have been caused by a new sort of radiation, being emitted by the gas discharge tube, and apparently capable of covering great distances. This strikingly powerful radiation were X-rays, today well-known to us.

(2) Rutherford's *discovery of the atomic nucleus*. In 1909 Rutherford was doing an experiment to examine the structure of the atom, which was designed following military rules: If you don't know what the object in front of you is, then you'd better shoot at it.[2] Rutherford used a radioactive material as ray gun, and shot a narrow ray of alpha-particles on a thin metal sheet. Behind this sheet, a fluorescent screen had been mounted, which, when an alpha-particle hit it, would document this with a microscopic flash of light. The big surprise, now, was that the alpha-particles were not just observed as being slightly redirected after hitting the sheet, but that their direction was changed altogether. Some were even repelled by the sheet, as if they had hit a solid wall. Rutherford later observed:

> It was quite the most incredible event that has ever happened to me in my life. It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you.[3]

[1] A comprehensive account of the development of the new in science may be found in: J. Mittelstrass, *Leonardo-Welt: Über Wissenschaft, Forschung und Verantwortung,* Frankfurt 1992, pp. 74-95 (I 4 'Die Wissenschaften und das Neue').

[2] See R.U. Sexl, *Was die Welt zusammenhält: Physik auf der Suche nach dem Bauplan der Natur*, Frankfurt and Berlin and Vienna 1984, p. 145.

[3] E. Rutherford, 'The Theory of Atomic Structure', in J. Needham and W. Pagel (eds.), *Background to Modern Science*. Ten Lectures at Cambridge Arranged by the History of Science Committee 1936 (Cambridge, 1938), p. 68.

Rutherford then tried to do justice to this ‘incredible event’ by developing his theory of the structure of the atomic nucleus. Some of the (positively charged) alpha-particles had directly hit the (also positively charged) atomic nuclei and were hence repelled by them in their initial direction of movement.

Such are the examples for the entirely unexpected discoveries. As examples of the discoveries which had been expected, to some extent, by the discoverer as well as by science in general, but which were novel in detail, let me present the following two: (1) Lavoisier’s *discovery of the conjoined composite nature of water*. Lavoisier had, since 1778, been looking for the oxide (the compound with oxygen) of hydrogen, discovered in 1766 by Cavendish, but had not made any tangible progress. It was only in 1781, when Cavendish noticed (and Lavoisier learned of this in 1783), that in an explosion of so-called detonating gas, hydrogen and oxygen could be permuted into their own weight in water, that Lavoisier inferred that the long searched-for hydrogen-oxide was really water itself, and that hence water was not an elementary substance, but a compound. So, what was not surprising here was the existence of a hydrogen-oxide; entirely surprising was the discovery that this oxide was the well-known substance water.

(2) Örstedt’s *discovery of electromagnetism*. In the natural philosophy of romanticism, all natural powers were thought of as expressions of a single and fundamental force, and this gave rise to the expectation that natural powers would have to be convertible into each other. In particular, this was thought to apply to electricity and magnetism. Under the influence of this idea, Örstedt searched for such a conversion and discovered, in 1820, more or less accidentally during a lecture, that a magnetic needle would be deflected by a wire lead conducting electricity. The connexion between electricity and magnetism was thus discovered. The novel and unexpected aspect of this effect, now consisted of the fact that the current would cause the magnetic needle to rotate. What had been looked for was a direct attraction or repulsion between the electric charge and the magnetic poles. What had not been expected was a circular magnetic field surrounding the conductor the current was flowing through. A mistaken background assumption had prevented the discovery of electromagnetism for a while; chance had to come to the rescue to lead to the right track.[4]

[4] See R.U. Sexl, *op. cit*., pp. 61-62. For a diverging account see J. Agassi, *Towards an Historiography of Science* (The Hague, 1963), pp. 67-74. Agassi assumes that the discovery was less accidental than traditionally thought.

Finally, two examples of discoveries which had been expected by the discoverer, but came as complete surprises to science: (1) Poisson's *white spot*. In about 1830 Fresnel presented the first complete version of a wave theory of light. The heart of this theory was the assumption that waves of light are of a transversal nature (that is, they occur perpendicularly to the direction of extension propagation). Poisson thought this theory to be wholly absurd, and to prove this, he deduced an apparently nonsensical consequence from it. According to Fresnel's theory, a white spot would have to occur at the midpoint of the shadow produced by a point-shaped source of light directed at a circular disc ('Poisson's white spot'). Arago then undertook the (to all appearances, redundant) labour to also demonstrate the falsity of this consequence experimentally. But, entirely surprisingly, he really did find the predicted spot. Although he was not the originator of the theory who predicted this novel effect, he could have done so, had he executed the deduction. In the case of a theoretical prediction of a novel effect, its empirical manifestation might surprise the general public, but not the theorist.

(2) Einstein's *prediction of the diversion deflection of light in the gravitational field*. Einstein had deduced, from the General Theory of Relativity, that light would be deflected from its initial trajectory by a certain angle by a gravitational field. In 1919, Eddington examined the position of stars, the light of which was passing the sun very closely. He then compared these positions with those of the same stars at the time these were more distant from the sun. Indeed, the predicted deflection was observed in the precise degree expected. This successful theoretical prediction was greeted with great surprise by the uninitiated. At the same time, the process of testing the General Theory of Relativity has been compared to the Catholic procedure of canonization ('November 6, 1919, the day on which Einstein was canonized').[5] The successful prediction of the deflection of light was one of the required miracles.

These examples teach that there is no simple way to arrive at the new in science, and that the diverse ways to arrive at the new are not simple. Furthermore, they are only rarely due to the strict following of scientific programmes, and this is why talk of scientific revolutions, which has again become popular with Thomas Kuhn's work in the history of science, is not that misguided. Indeed, scientific revolutions differ from political and

[5] A. Pais, *'Subtle is the Lord ...': The Science and the Life of Albert Einstein*, (Oxford and New York, 1982), p. 305.

social ones by having a much more varied potential for change, and, as a rule, for producing fewer losses, but, after all, this is not a disadvantage and not deplorable. Moreover, theories in science sometimes die more honourably than on the political or ideological stage. There, they often only come to an end with the biological end of their adherents – it is true, although this is not altogether unknown in science too.[6]

Science, in its search for the scientific new, does not just get driven by discoveries, accidental or non-accidental ones – one could express this by striding from truth to truth – but, surprisingly, also via errors, in a heuristic sense.[7] Let me give you a final example also for this heuristic fruitfulness of errors, Einstein's *derivation of the General Theory of Relativity*. This relied on principles which were partially motivated epistemologically. One of the principles pertinent and even essential for the formulation of the theory is *Mach's Principle*, as Einstein calls it. According to the ideas of Newton, there is one prime, indeed, truly immobile, system of reference, 'absolute space'; movements relative to absolute space are indicated by the presence of forces of inertia (for instance, centrifugal forces). Through such forces, accordingly, absolute space takes effect on the objects, while, at the same time, the objects are never able to disturb absolute space in its tranquil existence. Einstein considered the assumption of a uni-directional causation for inconsistence and instead assumed that the forces of inertia are explicable through the relative positions and movements of the bodies (an idea he attributed to Mach).

Mach's principle is not just one of the central motives for the development of the General Theory of Relativity, but also plays an important role in the process of formulating the theory. However, it turns out that the fully developed theory does not satisfy Mach's principle. That is, the theory allows space-time structures as physically possible, in which forces of inertia originate out of an overarching space-time that is independent of objects and their movements – even if this does not have the Newtonian shape of an absolutely immovable space. Furthermore, according to our current empirical knowledge, we may assume that in our universe, one such space-

[6] See M. Planck, *Wissenschaftliche Selbstbiographie* (Leipzig, 1948), p. 22 ('A new scientific truth does not usually become accepted by having its opponents convinced and having them declare their new conviction, but mostly by its opponents dying out, and having the new generation getting acquainted with the truth straightaway').

[7] Compare the more comprehensive account in J. Mittelstrass, *Die Häuser des Wissens: Wissenschaftstheoretische Studien* (Frankfurt, 1998), pp. 13-28 (I 1 'Vom Nutzen des Irrtums in der Wissenschaft').

time structure contradicting Mach's principle has been realised. Hence, Mach's principle would be false; following today's theoretical and empirical state of knowledge, our universe is equipped with a space-time structure that is in part independent of the mass-energy distribution bodies in the universe. Nevertheless, as I have explained, Mach's principle played an essential and probable indispensable role in the process of formulating the General Theory of Relativity. At least, the development of the theory would not have been possible along the path taken by Einstein without assuming this principle.

In other words, error too may play an essential role, not just not hindering scientific progress, but even furthering it. The 'game of science' (Popper) knows many successful moves; one of these is the truth of error.

DISCOVERY IN THE NEW COSMOLOGY OF COPERNICUS, KEPLER AND GALILEO

GEORGE V. COYNE, S.J.

INTRODUCTION

I would like to suggest three components contained in the notion of discovery: newness, an opening to the future and, in the case of astronomical discovery, a blending of theory and observation. Discovery means that something new comes to light and this generally happens suddenly and unexpectedly. While I would not exclude that one can plan and even predict what is to be discovered, this is generally not the case.

According to this description Copernicus truly discovered. With one stroke of genius he resolved what for centuries had been a most puzzling feature of Ptolemaic astronomy, namely, that the annual motion of the sun was coupled to the motion of any other planet. In all the planetary theories there was a decisive parameter equal to precisely one solar year. In Ptolemy this strange coincidence appears as a simple fact which is neither explained nor commented upon, and before Copernicus very few astronomers had paid attention to it. Now it found a simple explanation in Copernicus' rearrangement of the celestial orbits. The sun was placed in the center of the universe and the earth revolving around it as a planet with a period of one solar year. Although Aristarchus, for instance, had long ago proposed a heliocentric universe, Copernicus' discovery was truly new in that his insight resolved a centuries long puzzle and paved the way to the future, to Kepler and Galileo and eventually to Newton. He was truly a revolutionary.

For all of his life Kepler remained convinced that the secrets of nature can only be disclosed by mathematical methods. However, when, using only mathematics he failed to determine a circular orbit for Mars that would satisfy Tycho Brahe's and his own observations, he changed his method to an attempt to derive the motion of a planet from its cause. Assuming that the motive force comes from the sun and is inversely proportional to the distance from it, Aristotle's dynamical law of the proportionality of force and velocity showed immediately that the 'area velocity' of the planet is the same at both the apogee and the perigee. In the *Astronomia Nova* (1609)[1] Kepler made the assumption that it is constant all over the orbit. This is now known as Kepler's Second Law, a true discovery, since for the first time in the history of science, an intuited theoretical cause, i.e., a sun centered force, linked to observations had yielded a new result. This discovery plus that of the other two famous laws of Kepler looked to the future for a discovery of the ultimate cause. Kepler was surely one of the principal spans in the bridge from Copernicus to Newton.

Galileo was the first true observational astronomer[2] but he was also an experimentalist. In fact, it was precisely through his dedication as an experimentalist, and in particular through his studies on motion that he had come to have serious doubts about the Aristotelian concept of nature. What he sensed was lacking was a true physics. The world models inherited from the Greeks were purely geometrical and the geometry was based upon preconceived philosophical notions about the nature of objects in the universe: all objects had a natural place in the universe and consequently they had a natural motion. But there was no experimental justification for these preconceptions. They were simply based upon a philosophical idea of the degree of perfection of various objects. And then he came to know of the telescope. Did he have expectations that he would with the telescope find out for certain whether the world was Copernican? I expect not. His expectations were not that specific. He simply knew that the small instrument he had worked hard to perfect, if he had already convinced his patrons of its value for military purposes, was surely of some value for scientific purpos-

[1] *Astronomia Nova*, in M. Caspar (ed.), *Gesammelte Werke*, München, 1937, III.

[2] My claim that Galileo was the first true observational astronomer requires some justification. Galileo did not invent the telescope; he improved it for the precise purpose of astronomical observations. Nor, it seems, was he the first to use the telescope to observe the heavens. There is evidence that Thomas Digges of England, using a rudimentary reflecting telescope invented by his brother Leonard, saw myriads of stars about thirty years before Galileo's first observations. Even if this is true, the observations of Digges did not become known and had no scientific impact. Galileo not only observed; he intuited the great importance of his observations and he communicated them rapidly to the whole cultured world of his day. It is for that reason that I feel justified in naming him the first true observational astronomer.

es. That in itself, although it may seem trite to us, was a major discovery.[3] He truly discovered; he contributed by his telescopic observations the first new data about the universe in over 2,000 years. He was the first of many to follow. They are 'the people with long eyes'.[4]

Each of these three great figures was in his own unique way a discoverer. Let us examine now in more detail what unique contribution each made through his discovery to development of cosmology.[5]

Copernicus, the Revolutionary

Up until Copernicus arguments about cosmology were located within a mathematical and descriptive discourse on celestial phenomena without any significant reference to their immediate causes. In fact, because of that long historical antecedent, when in 1543 the great work *De revolutionibus orbium coelestium* (On the Revolutions of the Heavenly Orbs)[6] by Copernicus appeared its purpose was immediately misunderstood. It was provided with several prefaces, one of which was a letter from Cardinal Nicholas Schönberg written in 1536 CE and urging Copernicus to publish his new theory of the world. Another preface by Copernicus himself dedicated the work to Pope Paul III to whom the author explained his reasons for his drastic change of the traditional views on cosmology, hoping that the work might 'make some contribution also to the Church' by furthering the efforts to reform the Gregorian calendar, adding that *Mathemata mathematicis scribuntur*'.[7] This appeal to a certain autonomy of the mathematical

[3] It indeed was a major discovery to intuit the importance of the telescope for investigating the universe. In Note 2 I have remarked that Thomas Digges may have actually been the first to observe with a telescope but it appears that he did so in a rather perfunctory fashion and without an appreciation of its value for science, or at least he did not communicate that science to the world.

[4] This description of astronomers is due to the Tohono O'Odham Indians of the southwestern United States who, not otherwise having a word for astronomer in their language, constructed this phrase to describe those who came to build telescopes in collaboration with them on their sacred mountains.

[5] I am very much indebted to the lifelong research of Olaf Pedersen (deceased in 1997) for much of this historical survey. I am preparing the posthumous publication of his voluminous work, *The Two Books*, a brief presentation of which is given in *The Book of Nature*, Vatican City, Vatican Observatory Publications, 1992.

[6] Reprinted in the edition of Nicholas Copernicus, *Complete Works* published in several volumes by the Polish Academy of Science, Warsaw-Cracow, 1973.

[7] 'Mathematical matters are written for mathematicians,' Copernicus, *Complete Works*, 1978, II, 5.

discourse on nature made him forestall incompetent objections: 'Perhaps there will be babblers who claim to be judges of astronomy although completely ignorant of the subject and, badly distorting some passage of Scripture to their purpose, will dare to find fault with my understanding and censure it. I disregard them even to the extent of despising their criticism as unfounded'.[8] This premonition of coming disaster was perhaps the reason why the book contained also a third preface by a Lutheran Minister, Osiander, in which the serious nature of the Copernican reform was explained away as a purely abstract, mathematical game, just more of the same as throughout the previous centuries, that would not disturb the established order of the universe. It is the duty of the astronomer, it is said, to study the celestial motions and to 'conceive and devise the causes of these motions or hypotheses about them. Since he cannot in any way attain to the true causes, he will adopt whatever suppositions enable the motions to be computed correctly from the principles of geometry for the future as well as for the past. The present author [Copernicus] has performed both these duties excellently. For these hypotheses need not be true nor even probable. On the contrary, if they provide a calculus consistent with the observations, that alone is enough'.[9]

The very wording of this anonymous preface shows that it could not stem from Copernicus' own hand,[10] nor did it express his own view of the ontological status of his now hypothesis, although a first reading of the work might easily suggest that the opposite were true, since there is no more physics in the *De Revolutionibus* than in Ptolemy's *Almagest* which Copernicus tried to revise, removing from it some of its less satisfactory features. Among these were Ptolemy's use of circular motions that were not uniform with respect to their own centers. Copernicus got around this aesthetic difficulty by devising a new system of planetary epicycles that was at least as complex as Ptolemy's own. However, as we have seen in the introduction, Copernicus' great revolution was to solve the long-standing puzzle of the one solar year dependence of all former planetary theories with a simple rearrangement of the celestial orbits. The sun was placed in the center of the universe with the earth orbiting it. Consequently, any position of a star must change with this period according to the changing position of

[8] *Ibid*.

[9] Copernicus, *Complete Works*, 1978, II, xvi.

[10] In fact it was written by the Lutheran minister Andreas Osiander who saw the book through the press at Nuremberg. This was first discovered by Kepler.

the terrestrial observer. This solved the riddle of the Ptolemaic theories at the same time as it gave rise to the new question as to whether the fixed stars would also show an annual variation (or parallax) of a similar kind. An observational answer to this question would provide a deciding argument in favor of the Copernican system. Since no such parallax was discovered, Copernicus assumed, correctly so as we now know, that the sphere of the fixed stars, which was to him still the outer boundary of the universe, must be much further away than previously suspected.

Copernicus, contrary to Osiander and almost all cosmologist in the previous centuries, clearly realized that, even if various astronomical theories would save the same celestial phenomena, they might have different physical implications. This is obvious from his remark that it would be necessary to change the concept of gravity. Terrestrial gravitation could no longer be construed as a tendency to move towards the center of the universe, for the earth was no longer there. Instead he declared his belief 'that gravity is nothing but a natural desire, which the divine providence or the Creator of all things has implanted in parts, to gather as a unity and a whole, by combining in the form of a globe. This impulse is present, we may suppose, also in the sun, the moon, and the other brilliant planets, so that through its operation they remain in that spherical shape which they display'.[11] The only possible inference is that Copernicus believed his celestial kinematics to be a true account of the planetary system, even if, and this is the important point here, it was developed without consideration of the possible causes of the motions of its parts. Nonetheless, he also believed that this would lead him to true insights into reality as it had been shaped by the Creator. However, this passing reference to creation is practically the only point at which Copernicus referred to theological matters.

Kepler, the Bridge to the Future

In fact, it is not until Kepler appeared that a scientist of the first order seriously considered his research in the light of his religious faith. Being always a convinced Copernican he tried in his first work, the *Mysterium Cosmographicum* (1596) to calculate the relative radii of the planetary orbits from the geometrical properties of the five regular polyhedra.[12] This

[11] Copernicus, *De Revolutionibus*, I, 9; in *Complete Works*, 1978, II.

[12] See the figure in *Mysterium Cosmographicum*, in M. Caspar (ed.) *Gesammelte Werke*, München, 1938, I.

was an attempt to derive information on the universe from purely geometrical structures that were known a priori. As such it clearly belonged in the Platonic tradition. In a similar vein the ratio of the orbits of Jupiter and Saturn seemed to appear from a special geometrical construction in the plane to which Kepler was led by considerations of astrology in which he was always an ardent believer. However, the results did not answer to the standard of accuracy which Kepler demanded and he had to admit his failure. He also failed, as we have seen in the Introduction, by his sheer mathematical method to determine a circular orbit for Mars. When he changed his method from pure mathematical analysis to a search for the cause of the motions he discovered the law of areas and armed with this result, he returned to the problem of the shape of the orbit which he attacked by making no *a priori* assumptions. He succeeded after very much numerical work in showing that Mars moved in an ellipse.[13] The generalized assumption that all planetary orbits are elliptical expresses Kepler's First Law.

Kepler's Third Law appeared in a later work called *Harmonices Mundi* (1619) in a rather unexpected context. The book begins with a very long exposition of the mathematical theory of music in the tradition going back to the Pythagoreans. And so it has often been thought that the Third Law came to light by applying musical theory to Copernican astronomy, and it is possible that Kepler himself regarded his work as a kind of updated version of the 'Harmony of the Spheres'. Nevertheless, a closer examination of the text reveals that Kepler subjected the planetary periods of revolution 'T' and their mean distances from the sun 'a' to trial and error calculations and was fortunate enough to realize that the ratio T^2/a^3 is the same for all the planets.[14] So here again it really was an a posteriori method which led to the result.

No previous scientist had ever been able to carry this type of mathematical approach to nature to a similar perfection with such a methodological freedom and open-mindedness. More than anyone else it was Kepler who became the herald of a new era in which mathematical physics would go from strength to strength, and his works remain a source of wonder and admiration for all who have the patience to follow all the turnings of his mind and to repeat all his computations.

[13] *Astronomia Nova*, IV, 49, in M. Caspar (ed.) *Gesammelte Werke*, München, 1937, III, 367, ff.

[14] *Harmonices Mundi*, V, 3, in M. Caspar (ed.) *Gesammelte Werke*, München, 1940, VI, 302.

But Kepler was also unique in another way. Despite their mathematical character his works were written in a highly personal style which faithfully explained all the mistakes he had made and all the blind roads he had explored before the final result turned up. This makes them difficult reading; but it also provides a rare psychological insight into the hopes and fears of his research, revealing both his frequent despair and his final jubilation when the goal was reached and a new property of nature disclosed. Recapitulating his work in the *Mysterium Cosmographicum* he wrote: 'How intense was my pleasure at this discovery can never be explained in words! I no longer regretted the time wasted day and night. I was consumed by computing in order to see whether this idea would agree with the Copernican orbits, or if my joy would be carried away by the wind'.[15] This excitement was caused by an idea which failed and was in a way carried away by the wind. No wonder that a success made him even more jubilant as when he told about the discovery of the Third Law: 'Since the dawn eight months ago, since the broad daylight three months ago, and since a few days ago when the sun illuminated my wonderful speculations, nothing holds me back! I dare frankly to confess that I have stolen the Golden Vessels of the Egyptians to build a tabernacle for my God far from the bounds of Egypt [...]. The die is cast, and I am writing the book, to be read now or by posterity, it matters not! It can wait a century for a reader, as God Himself has waited 6000 years for a witness'.[16]

This text deserves a close reading. In particular the reference to the Golden Vessels is highly significant. As told in the story of the Exodus[17] these vessels were stolen or borrowed by the people of Israel when they left Egypt. They were sacred treasures, not of their own making, but acquired from a foreign country. This clearly reminded Kepler of the manner in which his most spectacular discoveries had emerged. Every time he had tried to impress a mathematical structure of his own choice upon nature he had failed. The laws he found were not of his own making. Success always came a posteriori and in unexpected ways. Thus the discovery of the Third Law convinced him that 'the whole nature of harmonies in the celestial movements does really exist, but not in the way I previously thought, but in

[15] *Mysterium Cosmographicum*, Praef., in M. Caspar (ed.) *Gesammelte Werke*, München, 1938, I, 13.

[16] *Harmonices Mundi*, in M. Caspar (ed.) *Gesammelte Werke*, München, 1940, VI, 290.

[17] *Exodus* 12, 35.

a completely different, yet absolutely perfect answer'.[18] His results came as delightful surprises, precisely because they were not really fruits of his own intellect. Despite all the mental energy he had spent unraveling them, they were actually stolen from a foreign country outside his own mind.

Kepler gave this insight a deeply religious interpretation, realizing an inner harmony between his scientific and his spiritual vocation. And in his last great exposition of astronomy the same idea was still alive and prominent in his thought about the study of astronomy: 'For it is precisely the universe which is that Book of Nature in which God the Creator has revealed and depicted his essence, and what he wills with man, in a wordless [*alogos*] script'.[19] This looks like a rather traditional instance of how the metaphor of the Book could be used with the backing of the Epistle to the Romans. But Kepler drew from it some personal consequences which shows how intense was his feeling of touching the Divine by means of his research. First it shed a revealing light upon his own self. Having undertaken to praise God in his works in the *Mysterium Cosmographicum*, he continued as if he suddenly realized the seriousness of what he had just put on paper: 'Therefore I am so stupefied that I must exclaim with Peter: "Go away from me for I am sinful man!"'.[20] And a couple of years later the same sense of being close to God in his work caused him to use a daring expression which had never before been heard in Christendom: 'Since we astronomers are Priests of the Most High God with respect to the Book of Nature, it behoves us that we do not aim at the glory of our own spirit, but above everything else at the Glory of God'.[21] Twenty years later he had found no reason to abandon this understanding of the vocation of the scientist. In a preface to the Emperor he explained that 'I understand myself as a priest of God the Creator, appointed by your Imperial Majesty'.[22]

This audacious notion of the scientist as a priest calls for two brief comments. Firstly, in Kepler it was not based on any sort of pantheism. He was not a priest of nature, but a priest of the Book of Nature; and the author of

[18] *Harmonices Mundi*, V, Proem., in M. Caspar (ed.) *Gesammelte Werke*, München, 1940, VI, 189.

[19] *Epitome Astronomiae Copernicanae* (1618) in M. Caspar (ed.) *Gesammelte Werke*, München, 1953, VII, 25.

[20] In the letter to Maestlin, 1595 October 3, quoted above.

[21] Letter to Herwath von Hohenburg, 1598 March 26, in M. Caspar (ed.) *Gesammelte Werke*, München, 1945, XIII, 193.

[22] *Epitome Astronomiae Copernicanae* (1618) in M. Caspar (ed.) *Gesammelte Werke*, München, 1953, VII, 9.

his book was a God who was not an immanent force in nature but a true Creator. He was even more; for to know him did not mean that man would see himself as a little spark of the Divine or an element of the great unity of everything. It meant that man saw himself as a sinner. Secondly, it is worth noticing that Kepler here took a significant step beyond Aristotle who had denied that any wisdom could be found in any kind of mathematics, and who had based his natural knowledge of God on the idea of causality in nature. Kepler was of a different opinion. The student of nature can realize the Glory of God even if he knows nothing about the causes of its phenomena. Kepler knew that he had found the three laws without really knowing the causes of planetary motion, even if he vaguely suspected it to be of a magnetic nature. And from a religious point of view it was enough to know that the relations or 'harmonies' of nature did really exist. This interpretation is supported by his brief meditation in the *Astronomia Nova* on King David the Psalmist who praised God for the wonders of the heavens even if he was 'far removed from speculating about physical causes, being completely at rest with the Greatness of God who made all this'.[23] Also for Kepler the mere 'givenness' of the constant relations in nature was the decisive point. Therefore, his intellectual conquests did not lead him to pride and vainglory, but to the humility of a man who was drawn out of himself to be at rest in that which was not his own.

Galileo, Eyes on the Universe

With Kepler the Book of Nature reached the summit of its metaphorical life as the vehicle of the self-understanding of a first rate scientist who was deeply committed to the Christian Faith. But with Galileo the Book of Nature was confronted with the Book of Scripture in a dramatic encounter which has ever since been regarded as one of the most decisive interactions between the world of science and the world of belief. Many polemicists have even taken it as the final proof of the alleged incompatibility of these two worlds and evidence of an essential enmity between the Catholic Church and the scientific attitude.[24] It became Galileo's fate to shake the

[23] *Astronomia Nova*, Dedic., in M. Caspar (ed.) *Gesammelte Werke*, München, 1937, III, 31.

[24] An excellent up-to-date study of the Galileo affair up until the most recent statements of John Paul II is: A. Fantoli, *Galileo: For Copernicanism and for the Church,* Vatican City, Vatican Observatory Publications 2002, Third English Edition; distributed by the University of Notre Dame Press.

foundations of the inherited discourse on nature. In this he was not alone, but in the course of the events he came to occupy a more conspicuous position than that of the other great pioneers of modern science. Copernicus had lived and worked almost unseen and unnoticed in a remote corner of Poland, and even if Kepler was the Imperial Mathematician his works were much too technical to draw the attention of more than a few experts. But, after many years of quiet work at Pisa and Padua, Galileo suddenly rode to European fame in 1610 when he published the first results of his epoch making astronomical observations with the telescope he had constructed (but not invented). All the world was amazed at the mountains on the moon, the innumerable fixed stars, the resolution of the Milky Way into separate stars, and the four satellites revolving around the planet Jupiter.[25] The framework of traditional cosmology had no room for such discoveries and would collapse under their weight.

During the very last year of what he himself described 'as the best [eighteen] years of his life'[26] spent at Padua Galileo first observed the heavens with a telescope. In order to appreciate the marvel and the true significance of those observations, we must appreciate the critical intellectual period through which Galileo himself was passing at the time those observations were made. As we have noted in the Introduction, Galileo was the first true observational astronomer but he was also an experimentalist. It is impressive, indeed, to visit the *Istituto e Museo di Storia della Scienza* in Florence where one sees the many broken lenses from Galileo's attempts to make ever better telescopes. He himself stated that 'of the more than 60 telescopes made at great effort and expense [in his home in Borgo de' Vignali in Padua] I have been able to choose only a very small number ... which are apt to show all of the observations'.[27] In that museum one also sees a display showing Galileo's application of the pendulum to a clock and his experiments with an inclined plane in search of the law of falling bodies. Before he pointed his finest telescope to the heavens he had done his best to show experimentally that there were no serious 'instrumental effects'. Again, in his own words: 'In so far as I can truthfully state it, during the infinite, or, better said, innumerable times that I have looked with this instru-

[25] *Sidereus Nuncius*, Venice, 1610, in *Edizione Nazionale delle Opere di Galileo*, reprinted 1968, Florence, Giunti Barbèra, A. Favaro (ed.) III, 55-96. English translation in Stillman Drake, *Discoveries and Opinions of Galileo*, New York, 1957, 21-58.

[26] A. Favaro, *op. cit*. In Note 25, vol. XVIII, 209.

[27] *Ibid*.

ment I have never noticed any variation in its functioning and, therefore, I see that it always functions in the same way'.[28]

Before we turn our gaze upon Galileo with his perfected telescope pointed to the heavens, I would like to attempt to recover his state of mind at that moment. This is admittedly a very tendentious thing to do, but I think it is important to attempt to do so for the sake of understanding what we might possibly mean by 'discovery'. He was nearing the end of a relatively long, tranquil period of teaching and research, during which he had come to question at its roots the orthodox view of the known physical universe. But he had as yet no solid physical basis upon which to construct a replacement view. He sensed a unity in what he experienced in the laboratory and what he saw in the heavens. For Galileo, the motion of falling bodies and the motion of the planets had something in common and geometrical explanations were not sufficient. Physics was required. But, in addition to his attachment to experiment and the sense for physics that derived from it, Galileo also nourished the idea that the true physical explanation of things must be simple in the richest meaning of that word. To be more specific, among several possible geometrical models the nature of the physical world would see to it that the simplest was the truest. But his view of the physical world was limited, although there was some expectation that, since with his telescope he had seen from Venice ships at sea at least ten times the distance at which they could be seen by the naked eye, he might go a bit beyond that limit. He was uncertain about many things in the heavens. He had seen an object suddenly appear as bright as Jupiter and then slowly disappear; he had been able to conclude that it must be in the realm of the fixed stars, but he could venture nothing about its nature.

Obviously not everything happened in the first hours or even the first nights of observing. The vault of the heavens is vast and varied. It is difficult to reconstruct in any detail the progress of Galileo's observations; but from October 1609 through January 1610 there is every indication that he was absorbed in his telescopic observations. From his correspondence we learn that he had spent 'the greater part of the winter nights under a peaceful open sky rather than in the warmth of his bedroom'.[29] They were obviously months of intense activity, not just at the telescope but also in his attempt to absorb and understand the significance of what he saw. At

[28] *Ibid*.
[29] *Ibid*.

times his emotional state breaks through in his correspondence. He makes a climatic statement in this regard in a letter of 20 January 1610, some weeks after his observations of the Medicean moons of Jupiter, when he states: ‘I am infinitely grateful to God who has deigned to choose me alone to be the first to observe such marvelous things which have lain hidden for all ages past’.[30] For Galileo these must have been the most exhilarating moments of his entire life.

But he will be very acute and intuitive when it comes to sensing the significance of his observations of the moon, of the phases of Venus, and, most of all, of the moons of Jupiter. The preconceptions of the Aristotelians were crumbling before his eyes. He had remained silent long enough, over a three month period, in his contemplations of the heavens. It was time to organize his thoughts and tell what he had seen and what he thought it meant. It was time to publish! It happened quickly. The date of publication of the *Sidereus Nuncius* can be put at 1 March 1610, less than two months after his discovery of Jupiter’s brightest moons and not more than five months after he had first pointed his telescope to the heavens. With this publication both science and the scientific view of the universe were forever changed, although Galileo would suffer much before this was realized. For the first time in over 2,000 years, new significant observational data had been put at the disposition of anyone who cared to think, not in abstract preconceptions but in obedience to what the universe had to say about itself. Modern science was aborning and the birth pangs were already being felt. We know all too well how much Galileo suffered in that birth process.

Galileo’s telescopic discoveries dramatically overturned the existing view of the universe. They looked to the future. Were there other centers of motion such as seen with Jupiter and its moons? Did other planets like Venus show phases and changes in their apparent sizes? And what to make of those myriads of stars concentrated in that belt which crosses the sky and is intertwined with bright and dark clouds? All of these were questions for the future. Although neither Galileo nor any of his contemporaries had a well developed comprehensive theory of the universe, Galileo clearly intuited that what he saw through his telescope was of profound significance. His discoveries were not limited to looking; they involved thinking. Henceforth no one could reasonably think about the universe in the tradition of Aristotle which had dominated thinking for over two millennia. A new theory was required.

[30] *Ibid*.

Did Galileo's telescopic discoveries prove the Copernican system? Did Galileo himself think that they had so proven?[31] There is no simple answer to these questions, since there is no simple definition of what one might mean by proof. Let us limit ourselves to asking whether, with all the information available to a contemporary of Galileo, it was more reasonable to consider the Earth as the center of the known universe or that there was some other center. The observation of at least one other center of motion, the clear evidence that at least some heavenly bodies were 'corrupt', measurements of the sun's rotation and the inclination of its axis to the ecliptic and most of all the immensity and density of the number of stars which populated the Milky Way left little doubt that the Earth could no longer be reasonably considered the center of it all.

But Galileo was also wise enough to know that not everyone could be easily convinced. In a letter to Benedetto Castelli he wrote: '... to convince the obstinate and those who care about nothing more than the vain applause of the most stupid and silly populace, the witness of the stars themselves would not be enough, even if they came down to the Earth to tell their own story'. While he could not bring the stars to Earth, he had, with his telescope, taken the Earth towards the stars and he would spend the rest of his life drawing out the significance of those discoveries.

SUMMARY

At the beginning of this essay I suggested three components contained in the notion of discovery: newness, an opening to the future and, in the case of astronomical discovery, a blending of theory and observation. Copernicus, Kepler and Galileo, according to these criteria, were all discoverers in the search for an understanding of the universe, but each in his own unique way. Copernicus was truly a revolutionary. With one stroke of genius by rearranging orbits in the solar system he resolved what for centuries had been a most puzzling feature of Ptolemaic astronomy, namely, that the annual motion of the sun was coupled to the motion of any other planet. Kepler was truly a bridge to the future. With his unique combination of an intuitive idea about the force of the sun and

[31] Historians debate endlessly as to when Galileo first became personally convinced of the correctness of Copernicanism. Judging from his statement of 'already for many years' and from other indications he must have certainly been leaning towards Copernicanism during the first years of his teaching at Pisa, which began in 1589.

a mathematical analysis of observations, he discovered the three laws of planetary motion named for him, thus leading the way to Newton and the law of universal gravity. Galileo can be rightly called the first 'person with long eyes'. He contributed by his telescopic observations the first new data about the universe in over 2,000 years. But he was also an experimentalist. He sensed a unity in what he experienced in the laboratory and what he saw in the heavens.

THE DISCOVERY OF EXTRASOLAR PLANETS

PIERRE LÉNA

In the fall of 1995, a question which had been haunting astronomers and philosophers for twenty-five centuries found its first clear answer, given by the undisputable observation of the first planet around a star other than the Sun, 42 light-years away: 51 Pegasi had a planetary companion, orbiting in 4.2 days and at least half as massive as Jupiter. The astronomers Michel Mayor and Didier Queloz, from the Geneva Observatory, had made this extraordinary discovery[1] at the Observatoire de Haute-Provence, in France, after ten years of efforts, concluding a quest which had begun 50 years earlier and opening a radically new era in astronomy.

Since this first observation, 155 exoplanets (also called *extrasolar planets*)[2,3] have been discovered around stars more or less similar to the Sun, with 120 cases where a single planet is known, and 13 cases of multiple systems formed of 2 or 3 planets (see Fig. 1, page 282). To take the month of September 2004 alone, 11 planets were published! The least massive planet known to date orbits around the star μ Arae, with a mass of 14 times (lower limit) the mass M of the Earth.

A Question Raised Twenty-five Centuries Ago

This discovery deserves careful attention for two reasons. First, contrary to many discoveries which seem to come unexpected or were hard-

[1] Mayor, M., Quéloz, D., 'A Jupiter-mass companion to a solar-type star', *Nature*, 378, 355 (1995).

[2] For a presentation of up-to-date discoveries in this field, see the remarkable site by Schneider, J. *Encyclopédie des planètes extrasolaires*, www.obspm.fr/encycl/f-encycl.html.

[3] Marcy, G., *et al.*, http://exoplanets.org.

ly guessed even a few decades before their outburst, the question of other worlds has been present since the dawn of philosophy and science in humanity. Raised implicitly or explicitly by anyone who contemplates the sky at night, it deals with the very place of man in the Universe, as it was admirably put forward by Blaise Pascal when he writes in his *Pensées*:

> Que l'homme contemple donc la nature entière dans sa haute et pleine majesté, qu'il éloigne sa vue des objets bas qui l'environnent... Mais si notre vue s'arrête là, que l'imagination passe outre; elle se lassera plutôt que de concevoir. Tout ce monde visible n'est qu'un trait imperceptible dans l'ample sein de la nature. Nulle idée n'en approche... Que l'homme, étant revenu à soi, considère ce qu'il est au prix de ce qui est; qu'il se regarde comme égaré dans ce coin détourné de la nature; et que de ce petit cachot où il se trouve, j'entends l'univers, il apprenne à estimer la terre, les royaumes, les villes et soi-même à son juste prix. Qu'est-ce qu'un homme dans l'infini?[4]

The second reason is the following: the quest for exoplanets has succeeded because of the intrication between a deep, long-lasting interrogation over centuries and the development of observational techniques and instruments in astronomy, an intrication which is outstanding in this case but far from being typical of discovery in general. As I have devoted a great deal of my professional life to the latter, it is for me a good motivation to present this contribution to the 2004 Session *Paths of Discovery* of the Pontifical Academy of Sciences.

Probably, the question was first explicitly raised by Democrites and restated by Epicure, who immediatly related it to the infinity of the world:

> Il n'y a donc rien qui empêche l'existence d'une infinité de mondes ... On doit admettre que dans tous les mondes, sans exception, il y a des animaux et tous les autres êtres que nous observons ...[5]

The theological implications, especially on the Incarnation, of such a speculative statement made it a matter of dispute in Christianity, with positive (Albert the Great, 1206-1280) and negative (Etienne Temple) views on its pertinence. It certainly reached its most dramatic climax in 1600 with the death sentence pronounced against Giordano Bruno. As

[4] Pascal, B. *Pensées*, Fragment 72 (classification Brunschvicg), Lattès, Paris, 1988.

[5] Epicure, *Lettre à Hérodote*.

Nicolas de Cues (1401-1464) before him, Bruno refers to the omnipotence of God when considering the likelihood of other 'worlds':

> [Dieu] ne se glorifie pas dans un seul, mais dans d'innombrables Soleils, non pas en une seule Terre et un monde, mais en mille de mille, que dis-je? En une infinité de mondes.[6]

Certainly Bruno was a philosopher and a theologian, not a scientist as was Galileo: even so, his rehabilitation, as one prominent 'carrier in history' of an idea which has recently proven so fruitful, could be a recognition, by the Catholic Church, of the sinuous paths of discovery.

Emergence of a Scientific Problem

During the 19th and first half of the 20th centuries, the physical theories and the progress of astronomical observations helped to transform the previous speculations into a real scientific problem. Newton's theory of gravitation had proven its power to describe the stellar motions, especially in double or multiple systems of stars, which are extremely frequent in our Galaxy. It is indeed the gravitational perturbations of the planets in the solar system which led to the discovery of Neptune (LeVerrier 1846) and Pluto (Lovell 1936). It soon appeared that the discovery of exoplanets could only be indirect, and would result from tiny perturbations eventually measured on the star itself, i.e. on the change of its position on a sky photograph with respect to background 'fixed' stars appearing in its neighbourhood. Indeed, observing directly an utterly faint planet around a star with the telescope remained impossible: a companion B, only a thousand times fainter than the bright star Sirius A, was detected in 1862 by Alvan Clark on a photograph, but had previously (1844) been proposed by Friedrich Bessel on the basis of perturbations observed on Sirius' motion. It is interesting to observe that in the cases of Neptune or Pluto, the qualification of *discoverer* may refer to the person who predicted the existence and position of the object, or to the one who provided an observation proving directly an image or indirectly its presence. Arago said of LeVerrier:

> In the eyes of all impartial men, this discovery will remain one of the most magnificent triumphs of theoretical astronomy ...

[6] Bruno, G., *De l'infini, de l'univers et des mondes*, trans. J.-P. Cavaillé, Belles-Lettres, Paris, 1995.

Progress of Observations

By the mid-20th century, the performances of telescopes, associated with photography, began to make possible the detection of tiny perturbations in stellar motions, despite the continuous trembling of stellar images due to the presence of the Earth's atmosphere, which reduces the accuracy of astrometry and was already well-described by Newton.[7] From now on, the eventual *discovery* of an exoplanet would result from a *search* of these hypothetical objects, on the basis of quantitative predictions of their potential effects. Several astronomers (Strand in Philadelphia on the star 61 Cygni, Reuyl and Holmberg in Virginia on 70 Ophiuchi, Van de Kamp in Philadelphia on the Barnard star) did interpret their astrometric measurements as the proof of a planet in orbit. Since by then the well-established stellar theory allowed to establish the mass of stars, the mass of the assumed perturbator could then be deduced from the amplitude of the perturbation and its classification as a planet result from a comparison with Jupiter or Saturn's masses. Unfortunately for their authors, none of these claimed detections of planets were confirmed, and the Barnard candidate was definitively proclamed inexistant[8] in 1973. Would the classical astrometry method ever succeed in detecting planets?

But another question lays in the background of the search for exoplanets: how confident can one be that they simply exist elsewhere? could the Solar system be an absolute unique phenomenon in the Galaxy? – Detecting planetary systems in other galaxies was, and still is, entirely out of reach. Modern astronomers are all Copernicans in a generalized manner, in the sense that they refuse to give any privileged character to our present position as observers of the universe: in this respect, there *ought* to be other planetary systems, a statement which is not so far, although comforted by centuries of science, from the position of Epicure! But, in addition to this act of belief, understanding the history and formation of our solar system made great progress since the formulation of the *nébuleuse primitive* hypothesis by Pierre Simon de Laplace[9] in 1796. In 1902, Sir James Jeans proposed a scenario of gravitational accretion, which along with the formation of a star, would lead to the formation of a disc of gas and dust, and later to orbiting

[7] Newton, I., in *Optics*.

[8] Gatewood, G., Eichhorn, H., 'An unsuccessful search for a planetary companion of Barnard's star (BD +4 3561)', *Astron. J.*, 78, 769-776 (1973).

[9] Pierre-Simon de Laplace, *Exposition du système du monde*, 1st edn., Paris, 1976.

planets, a scenario which is now fully recognized in the light of precise simulations (see Fig. 2 and 3, pages 283-4, ref.[10]). This scenario would remain hypothetical if observation had not progressively confirmed it: in the last decade, tens of protoplanetary discs have been observed by the *Hubble Space Telescope* or ground based telescopes, giving a firm basis for the search of their last stage of evolution, namely the formation of planets.

Modern Times and Discoveries

The discovery of exoplanets could obviously rely on a small indirect effect on their parent star, but which effect should one detect, and with which instrument? During the period 1975-1990, the techniques of optical astronomy were making immense progresses. Because the progress of observation, which in astronomy is the source of discovery, required better images and better spectra, all the resources of optics, helped by the newborn and powerful computer capability, were explored by astronomers. It is important to point out here the weight of tradition and human models. In France for example, it is a long chain of physicists or opticians, who all showed an interest in astronomy, which led to the success of the last generation: there is a continuity of methods, traditions and open-minded exploration of the 'virtually impossible' from Fresnel, Biot, Fizeau, Chrétien, Lyot, Françon, Jacquinot, Maréchal, Connes, to finally Labeyrie who recreated optical interferometry,[11] abandoned since the prowess of Albert Michelson (1920), the first one to have directly measured stellar diameters. Similarly, it is in these years (1989) that adaptive optics emerged, beating for the first time in history the trembling of stars observed by ground-based instruments at visible or infrared wavelengths, and allowing to conceive and build new generations of giant telescopes, such as the European *Very Large Telescope* in Chile or the *Keck Telescopes* in Hawaii.

In 1970, a rudimentary but incredibly powerful spectrograph was built in Cambridge (UK) by Roger Griffin, exploiting ideas of multiplexing the light, which were put forward in the optics community at this time, both in England and in France: instead of measuring one spectral line among the ten thousands present in a stellar spectrum, one cumu-

[10] Artimovicz, P., www.astro.su.se/~pawel/.

[11] Labeyrie, A., 'Interference fringes obtained on Vega with two optical telescopes', *Ap.J.*, 196, L71-L75 (1975).

lates all the information common to these lines, improving enormously – by a factor of 10^3 – the sensitivity of the instrument to Doppler shifts. It is exactly what the people looking for variations of stellar velocities in multiple systems needed, and this did not escape the young Michel Mayor, who was nurtured by the long tradition of precise radial velocities measurements carried by the Observatoire de Genève, where he worked, and by the Observatoire de Marseille where a bright optician named André Baranne was conceiving new instruments. The first spectrograph they built entered operation in 1977, and was focused on the search for low mass stars in multiple systems.

In fact, stimulated by David Latham (Cambridge, USA) who was working along the same tracks, they jointly discovered with him in 1988 the object[12] HD114762, which was finally classified as a brown dwarf, i.e. an object of intermediate mass between a star and a planet: this result, coming after several false hopes of previous detections of these objects, was to lead to a list, in 2004, of several tens of brown dwarfs. By the end of the 1980s, the theory of star formation had placed fairly strict mass limits on the possible ignition of nuclear reactions in the core of an object to produce a real star (above 0.08 M_{sun}) or briefly on its surface to produce a brown dwarf (between 13 $M_{Jupiter}$ and 0.08 M_{sun}= 80 $M_{Jupiter}$), a mass of 13 $M_{Jupiter}$ thus defining the maximum mass of a planet. The brown dwarf quest is in itself another fascinating discovery story, which is connected to, but distinct from, the search for exoplanets, given the proximity of masses and detection techniques. The mass of HD114762 seems to make it a brown dwarf and not a planet, but doubt remains permitted.

At this point, Mayor comments that he became 'ouvertement un chasseur de planètes'[13] and a second generation of instruments was built, with Swiss accuracy, with all the available technology and the sensitivity required to detect the effect of a jovian planet on the parent star: this sensitivity can simply be expressed as the capability to measure, in the radial velocity of the star, temporal variations of the order of a few metres per second (a Doppler-Fizeau relative effect of roughly 10^{-8}). His team is not alone, as in California with Geoffrey Marcy and Paul Butler, as well as in Canada with Gordon Walker and Bruce Campbell, two other groups are on the

[12] Latham, D.W., Stefanik, R.P., Mazeh, T., Mayor, M., Burki, G., 'The unseen companion of HD114762 – A probable brown dwarf', *Nature*, 389, 38 (1989).

[13] Mayor, M., Frei, P.-Y., *Les nouveaux mondes du cosmos*, Seuil, Paris, 2001.

track: the potential discovery is nearly on hand, the technology is readily available, it all becomes a matter of tenacity, good weather... and luck.

The search made by these astronomers focused on the periodic fluctuations of the stars radial *velocity*, which, as a direct consequence of Kepler laws, is apparently caused by an orbiting planet and would allow the determination of its period and mass. But in the late 1970s, the old astrometric method, based on the measurement of periodic fluctuations in the star's *position*, was becoming possible at the requested accuracy (a fraction of a millisecond of an angle) thanks to the new but difficult technique of optical interferometry, reborn in 1975 and slowly developing. The astronomer Michael Shao, from the *Jet Propulsion Laboratory* in United States – a leading place in the exploration of the Solar system since the *Voyager* missions to Jupiter and Saturn in the 1970s and 1980s –, developed, in the 1980s on the historical Mt. Palomar in California, a remarkable astrometric interferometer to measure the orbit of close binary stars. He succeeded, but he soon realized that this indirect detection technique would only be really successful if carried in space by unmanned telescopes unaffected by atmospheric perturbations, and he initiated long-term projects in this direction, yet to come (see Fig. 4, page 284).

The odds were therefore in favor of the spectroscopists, and indeed, always needed at some point for a discovery to happen, was with the Geneva group. Luck seemed even more needed after an unsuccessful search, published in 1994 by the two other active groups in America, who found no companions on their sample of 20 stars: they were indeed looking for long orbital periods (years) as expected for objects comparable to Jupiter in mass a mass set by the limited sensitivity of the instruments – and far away from the star, where many theorists predicted the massive planets ought to be.

What follows is Mayor's description of the final run. In November 1994, he and his young collaborator Didier Quéloz hinted at an oscillation on one of their reference stars, 51 Pegasi. Having eliminated instrumental effects, they remained puzzled by the measured period of 4.2 days, which, given the known mass of this star, led to mass lower limit of half a Jupiter and an orbit situated at 0.05 astronomical units, i.e. extremely close to the star, a distance at which (almost) no one would imagine such a large planet to be present. At Saint-Michel-de-Haute-Provence Observatory, the star is no longer observable until July 1995, and they spend their time establishing the ephemerids of the planet in order to *predict* where it should be when it will be reobserved. This is an exact reminder of Le Verrier predicting, from celestial mechanics, the position where Galle in Berlin would observe Neptune. On July 6, 1995, the observation confirms precisely the prediction (see Fig. 5, page 285). The two astronomers prepare their publication in *Nature*, but in the meantime try to assess the physical possibility for an orbit to remain stable with such a close periastron, where tidal effects could quickly destroy the planet: their American colleague Adams Burrows, a specialist, is consulted and accepts with fair play to run his computer code without asking questions. Fortunately, his verdict does not rule out the stability and the discovery is announced and applauded at a scientific meeting in October in Florence. The discovery not only confirmed by their competitor Geoffroy Marcy at Lick Observatory, but the Lick group reviewed the data and searched them for short period planets, for which indeed they were not looking: they found two other planets (orbiting the stars 70 Virginis and 47 Ursae Majoris).

The entire astronomical community would soon understand that an entirely new field was open, and was forced to think over its prejudices against massive planets being so close to the star. In fact, one soon realized that a far-reaching analysis had been published[14] as early as 1986, analysis which could have led to focus the search on orbital periods of the order of days or weeks, rather than of years or decades. This analysis was showing that massive planets can indeed only be formed in the protoplanetary disc at a distance of several astronomical units of the star, such as Jupiter and Saturn's orbits, but can migrate inwards because of interactions of the young planet with the remaining gaseous disc.

Before drawing general conclusions from this splendid adventure, it is interesting to discuss briefly what the word *discovery* exactly means here, as the only proof at this moment was only indirect, i.e. the gravitational action of the planet on the star. No one had yet *seen* this planet, as Galle saw Neptune through his telescope. In 2000, the astronomer Henry, looking at the by then detected planets, found that one of them was regularly passing in front of its star because of the inclination of its orbit with respect to the line of sight from Earth. He observed a measurable 2% decrease of the star's brightness: not only this measurement gave a reasonable density of the planet, but it was adding another indirect proof of its presence. Five more such *transits* have since been detected.

Although in 2005 no one doubts the reality of the 155 discovered exoplanets, a real image of an exoplanet, as the one Galle obtained of Neptune, is much wanted, but indeed extremely difficult to obtain, as the planet is

[14] Lin, D.N.C., Papaloizou, J., 'On the tidal interaction between protoplanets and the protoplanetary disk. III – Orbital migration of protoplanets', *Ap.J.*, 309, 846 (1986).

buried into the much more intense light of the star. The detection of actual photons from an exoplanet would open the way to spectroscopy, i.e. to a detailed analysis of its composition, its atmosphere, etc.

Here comes another story, involving again a radical technological breakthrough, in which I was fortunate to play a direct role during the period 1981-1998: namely the advent of adaptive optics[15] to beat the deleterious effects of the Earth's atmosphere on astronomical images, which I mentioned above. To make a long story short, ground-based telescopes can be equipped with an active deformable mirror, controlled by a computer, which almost exactly compensates the effects of the atmosphere and restores the resolution capability the telescope would have if placed in space. This enormously helps the detection of faint stellar companions and the system we built for the European telescope *Yepun* (one of the four VLT telescopes) allowed Anne-Marie Lagrange and Gaël Chauvin to publish[16] in September 2004 the (likely) first *direct detection* of an object of planetary mass (approximately 5 M_J), orbiting around the brown dwarf 2M1207 (see Fig. 6, page 285): some caution is still exercized by the authors as absolute proof, eliminating the unlikely chance of a line-of sight coincidence with a background object, will only be given when the motion of the planet is directly observed.[17]

In this story of exoplanets, one important step is missing, which I should have emphasized in due time: it is the totally unexpected discovery of planetary mass objects around pulsars in 1992, with an entirely different method. Pulsars are neutron stars, left over after the explosion of a supernova: extremely compact and carrying most of the angular momentum of the parent star, they rotate as fast as 1 kHz, and are detected by their modulated radio or optical emission at this frequency. Radiofrequency measurements can be extremely precise (down to 10^{-12}) in relative accuracy, a performance which incidentally led to the first observational proof of gravitational waves, as predicted by general relativity. Because of the violent events which led to the formation of a pulsar, no one really believed that eventual planets existing around the parent star could survive the explosion, until successive hints of detections as early as 1970, only 3 years after

[15] Rousset, G., Fontanella, J.C., Kern, P., Gigan, P., Rigaut, F., Léna, P., *et al.*, 'First diffraction-limited astronomical images with adaptive optics', *Astron. & Astrophys.*, 230, L29-32 (1990).

[16] Chauvin G., Lagrange A.-M., Dumas C., Zuckerman B., Mouillet D., Song I., Beuzit J.-L. & Lowrance P., 'A giant planet candidate near a young brown dwarf', *Astron. & Astrophys.*, 425, L29 (2004).

[17] In 2005, the proof was fully given of the physical association of both objects, and the planet confirmed.

the discovery of the pulsars, although disproven later, stimulated theorists to think over the problem.

Any massive body orbiting around a pulsar will affect periodically the measured frequency, and become detectable. In 1992, the young Polish astronomer Alexander Wolszczan published the discovery of two planets, a third one being later added, orbiting around the pulsar PSR 1257+12: their masses are respectively 0.09, 3.9 and 4.3 times the mass of the Earth.[18] They are indeed, and will probably remain for some time, the least massive single objects ever detected at this distance from us: there is no dispute about the fact they are most likely the residual of a planetary system which survived the explosion of the supernova. Yet, they were and still are quite apart in the quest for other planetary systems, like ours, as the extremely strong radiation (γ and X rays) emitted by the pulsar does not leave any possibility for these objects to evolve towards complexity, as our system did. Despite many efforts, the list has not much increased in the ten last years, as it only counts another case as for 2004.

Conclusion

The story of exoplanets is in its infancy, and we should learn more in the coming years. As usual, the discovery is opening a kind of Pandora box, where many old questions suddenly become revitalized, the question of possible life emergence and evolution in *other worlds*, to use Epicure's expression, being the central one in the years to come.

The path of discovery has been long but quite straight. The initial question was fairly simple, as possibly the one which led to the concept of atom. The progress became significant by the mid-1900s, only when the observing tools reached the level of accuracy and sensibility required by the expected effects of an exoplanet. But to get an order of magnitude of these effects, a fairly large body of theory and extrapolation from our own solar system was needed at the same moment, and accompanied every step towards the final discovery. On the other hand, some of these predictions were taken too seriously and led to question the discovery, when the first exoplanet was found at such an odd close distance to its star.

[18] Wolszczan, A. & Frail, D., 'A Planetary System around the Millisecond Pulsar PSR1257+12', *Nature*, 255, 145 (1992).

One should note that this major discovery, opening a scientific revolution, is not the result of a radical change of paradigm, as its foundation has been laid down for centuries. Although the detection of the pulsar planets in 1992, for which no one was looking for, or the extreme proximity of the 51 Peg planet to its star, were both entirely unexpected, they could quickly be understood without a major revision of theories or models.

There is here an interesting point, namely the connection between discovery and *threshold* effects. Clearly Mayor's result came at a time where technology had suddenly reached the level or sensitivity needed for the detection. This may seem a self-fulfilling affirmation, but one may point out that several other paths with different instruments (e.g. astrometric measurements on the parent star, or transits measured from satellites), could as well lead to a similar discovery. Another threshold lay in the statistics involved:[19] the discovery could only result from a systematic and lengthy search on many stars... but no one knew a priori the size of a relevant sample. Walker, one of Mayor's competitors, worked unsuccessfully on a sample of 23 stars: with a now known probability of 4 to 5% for these stars to have a massive planet but studyng double stars, his chances to find one where dim. Mayor, who initially was not looking for planets, worked on a sample of hundreds of stars.

The long chain of astronomers who carried the issue and invented instruments by drawing on all the resources of optics available during their time has certainly been a key for success, and will likely remain so in the future. This continuity of efforts, the kind of *rage* to reach a result which seemed vanishing for so long, illustrates again the words of Bernard de Chartres[20] (†1130), quoted by John of Salisbury:

> Bernard de Chartres disait que nous sommes comme des nains juchés sur des épaules de géants [les Anciens], de telle sorte que nous puissions voir plus de choses et de plus éloignées que n'en voyaient ces derniers. Et cela, non point parce que notre vue serait puissante ou notre taille avantageuse, mais parce que nous sommes portés et exhaussés par la haute stature des géants.

In this respect, the final discoverer has the immense merit of the last step, of the tenacity to reach it, of the ability to seize the chance which is always needed, but is the heir of a long story.

[19] As pointed out by Jean Schneider, with whom the author had fruitful exchanges.

[20] Newton, to which this metaphor is often but wrongly attributed, used it in a letter to Robert Hooke in 1676, where he wrote: 'If I have seen farther than others, it is because I was standing on the shoulder of giants'.

GOING FROM QUARKS TO GALAXIES: TWO FINDINGS

RUDOLF MURADIAN

Introduction and Motivation

This talk is a brief summary of applications of power scaling laws in two different settings, in the micro- and the macro-world. The power scaling laws are widely used not only in physics, but also in life sciences, chemistry, and engineering. The remarkable scaling laws could be obtained by a combination of a few simple dimensional considerations with specific physical conjectures. The basis of method was presented for the first time by Galileo Galilei. Scaling ideas of Galileo it was later elaborated by Fourier, Lord Rayleigh and others and today is widely used in physics under the name of *Dimensional Analysis* [1]. The 2004 Nobel prize in Physics was awarded for revealing the 'asymptotic freedom' and 'infrared confinement' in QCD (Quantum Chromodynamics) by means of the renormalization group theory, which can be considered as the modern method for dealing with Galileo scaling.

The first part of this talk is based on common work with V. Matveev and A. Tavkhelidze [2] on the Dimensional Quark Counting Rules, later confirmed within the framework of QCD by G. Lepage and S. Brodsky [3].

The second part is dedicated to the suggestion of *spin-mass scaling rules in astrophysics* [8-12]. It is shown that these rules can be expressed by fundamental constants only without any adjustable parameter. A surprising resemblance between *angular momentum & mass* dependence of elementary particles and cosmic objects demonstrates remarkable unity of Nature on a huge scale. A probable solution of the old and most difficult problem of astrophysics – the problem of the origin of the rotation of planets, stars, galaxies and their systems – is proposed.

1. Scaling Laws and Quark Structure of Hadrons

A rise in interest in scaling laws in hadron physics was generated by the discovery of Bjorken scaling [4] and its generalization for lepton-hadron deep inelastic processes [5, 6]. This development was finally accomplished by the discovery of *Dimensional Quark Counting Rules* in exclusive processes [2] in 1973, before the development of QCD.

It is well known that, for understanding the structure of matter at smaller and smaller distances, it is necessary to collide particles with larger and larger momentum transfer. Transverse momentum $P_{\perp} \approx 10\, GeV/c$ corresponds to resolution of details at the distance of the order of $10^{-15}\, cm$.

For an exclusive two-body reaction $ab \rightarrow cd$ at high momentum $p_{\perp} \approx 10\, Gev/c$ transfer and fixed large scattering angle the Quark Counting Rule predict:

$$\frac{d\sigma^{ab \rightarrow cd}(s,t)}{dt} \sim \frac{1}{s^{n-2}} f\left(\frac{t}{s}\right) \tag{1.1}$$

where s and t are Mandelstam variables, s is the square of the total energy in the center-of-mass frame and t is the momentum transfer squared, $n = n_a + n_b + n_c + n_d$ is total number of constituents and n_i $i = a,b,c,d$ is the number of quarks in $1-$th hadron. The most interesting feature of the predicted exclusive cross-section is that it falls slowly, not as Gaussian or exponent, but as inverse power of, and the exponent of this power is directly related to the total number of quarks in scattering hadrons. This rule was derived in 1973 [2] and later confirmed experimentally in many exclusive measurements [7] and theoretically in the framework of perturbative QCD.

In [2] the asymptotic power law for electromagnetic form factors in particle and nuclear physics was also established. For hadronic or nuclear objects with constituents the corresponding form factor behaves as

$$F_n(t) \sim \frac{1}{t^{n-1}} \tag{1.2}$$

The amazing examples for agreement with the experiment are provided by the study of the form factors of pions, nucleons and light nuclei

$$F_{\pi}(t) \sim \frac{1}{t}, \; F_N(t) \sim \frac{1}{t^2}, \; F_d(t) \sim \frac{1}{t^5}, \; F_{He^3}(t) \sim \frac{1}{t^8}, \; F_{He^4}(t) \sim \frac{1}{t^{11}}. \tag{1.3}$$

2. Spin-Mass Scaling Laws and Origin of the Universe

At all scales of the Universe, from tiny elementary particles to huge celestial bodies, we observe rotating objects. Understanding the Universe is impossible without understanding the source of the spin or angular momentum of cosmic bodies. The Universe is filled with rotation: asteroids, planets and their moons, stars, interstellar clouds of gas, globular clusters, galaxies and their clusters rotate around a central axis, and everything orbits around everything else in a hierarchical manner (moons around their planet, planets around their star, stars around the center of their galaxy or globular cluster, galaxies around the center of their galaxy cluster). Spin or angular momentum is a conserved quantity: the total amount of rotation in the whole Universe must be constant. Rotation cannot just appear or disappear, but is innate, earliest, basic motion. When and how was the angular momentum acquired by celestial bodies? Can the rotation serve as the Rosetta Stone of Astrophysics? The problem of the origin of rotation in stars, galaxies and clusters is an open problem (*enfant terrible*) of astrophysics and cosmology. Here we will outline new insights into the origin of cosmic rotation, provided by the application of the Regge-Chew-Frautschi paradigm in astrophysical context.

Much of what we present here is based on [8]. In elementary particle physics after the works of T. Regge, G. Chew and S. Frautschi it has become clear that the spin J and mass m of hadrons are not independent quantities but are connected by a simple scaling relation

$$J = \hbar \left(\frac{m}{m_p} \right)^{2} \tag{2.1}$$

where $\hbar = 1.055 \times 10^{-34}\, J{\cdot}s$ is Planck's constant and $m_p = 1.673 \times 10^{-27}\, kg$ is the proton mass. This formula is well satisfied by experimental data obtained in high energy physics laboratories. The present author suggested following the extension of (2.1), using dimensional analysis and similarity considerations:

$$J = \hbar \left(\frac{m}{m_p} \right)^{1+1/n} \tag{2.2}$$

The number $n = 1, 2, 3$ in exponent takes the integral value n characterizing the spatial dimensionality (shape) of a spinning object. The choice

$n=2$ for galaxies, their clusters and superclusters, and $n=3$ for planets and stars are in brilliant agreement with the observations:

$$n=1: \qquad J = \hbar\left(\frac{m}{m_p}\right)^{2}, \qquad \textit{string (hadrons)} \tag{2.3}$$

$$n=2: \qquad J = \hbar\left(\frac{m}{m_p}\right)^{3/2}, \qquad \textit{disk (galaxies)} \tag{2.4}$$

$$n=3: \qquad J = \hbar\left(\frac{m}{m_p}\right)^{4/3}, \qquad \textit{ball (stars)} \tag{2.5}$$

These relations represent a surprising resemblance between *spin & mass* dependence of elementary particles and cosmic objects. It seems it is *the first time in the history of astronomy and physics that the spin of celestial bodies has been predicted from known mass using fundamental constants only.*

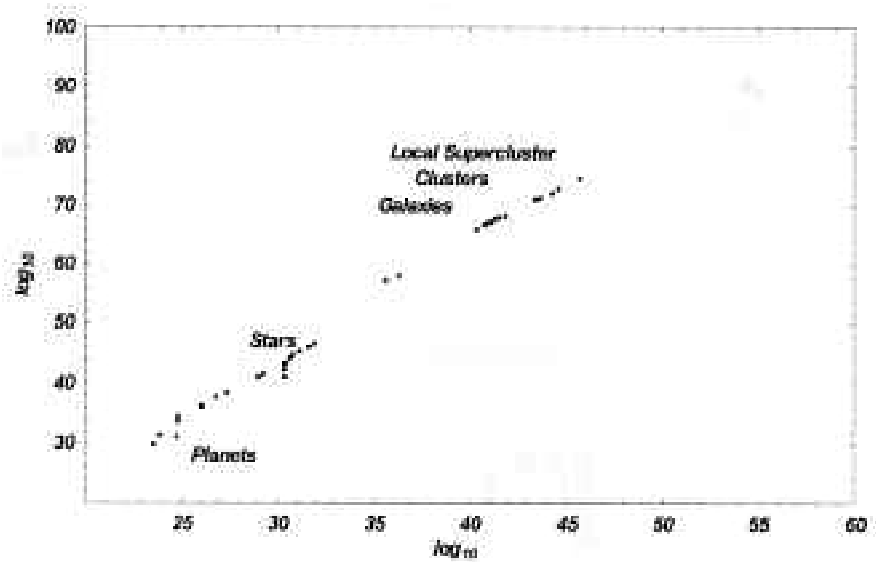


Fig. 1. Observational data on spins versus masses for diverse celestial objects. SI units are used.

For galaxies, their clusters and perhaps for the Universe itself Fig. 1 represents observational data on the $\log_{10}-\log_{10}$ plot for a whole spectrum of astronomical objects (see [11] and references therein). Fig. 2 is a pure theoretical construct and presents three straight lines

$$J = \hbar\left(\frac{m}{m_p}\right)^{4/3} = 53.114 \times m^{4/3} \tag{2.6}$$

$$J = \hbar\left(\frac{m}{m_p}\right)^{3/2} = 1.5412\times 10^{6} \times m^{3/2} \tag{2.7}$$

$$J_{Kerr} = \frac{Gm^2}{c} = 2.226\times 10^{-19} \times m^2 \tag{2.8}$$

where J_{Kerr} is gravitational Kerr spin-mass relation for maximally rotating black hole.

The coordinates of the intersection points are

$$m_{Eddington} = m_p\left(\frac{\hbar c}{Gm_p^2}\right)^2, \quad J_{Eddington} = \hbar\left(\frac{\hbar c}{Gm_p^2}\right)^3 \tag{2.9}$$

$$m_{Ch} = m_p\left(\frac{\hbar c}{Gm_p^2}\right)^{3/2}, \quad J_{Ch} = \hbar\left(\frac{\hbar c}{Gm_p^2}\right)^2 \tag{2.10}$$

These points can be named as

$$Eddington\ point \Rightarrow \{m_{Eddington}, J_{Eddington}\} \tag{2.11}$$

$$Chandrasekhar\ point \Rightarrow \{m_{Ch}, J_{Ch}\} \tag{2.12}$$

The *Eddington point* corresponds to the crossover of the Regge trajectory for disk-like objects with Kerr angular momentum. In the same manner the *Chandrasekhar point* corresponds to the crossover of the Regge trajectory for the ball with J_{Kerr}.

Fig. 3 represents an additive sum of Fig.1 and Fig. 2 and, in some sense, is a generalized Chew-Frautshi plot in astrophysics.

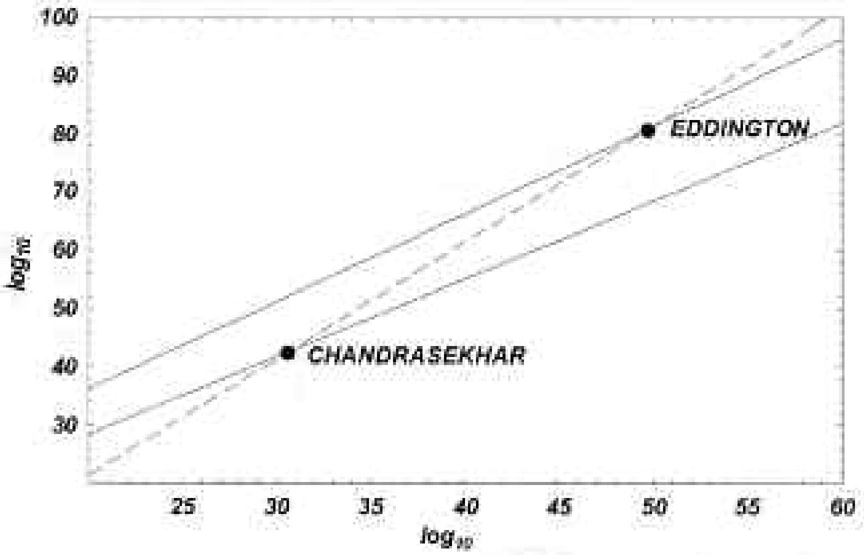


Fig. 2. Three theoretical straight lines are shown on the log-log plane.
a). The lower uninterrupted line

$$\log J = 1.725 + \frac{4}{3}\log m,$$

is a logarithmic representation of the formula (2.6) for three-dimensional ball-like objects.
b). The upper uninterrupted line

$$\log J = 6.176 + \frac{3}{2}\log m,$$

corresponds to the formula (2.7) for two-dimensional disk-like objects.
c). The dashed line corresponds to the Kerr angular momentum (2.8)

$$\log J = -18.651 + 2\log m$$

In his Nobel lecture Chandrasekhar put the question: ‘Why are stars as they are?’ and answered that it is because the following combination provides a correct measure of stellar masses:

$$m_n \left(\frac{hc}{Gm_p^2} \right)^{3/2}$$

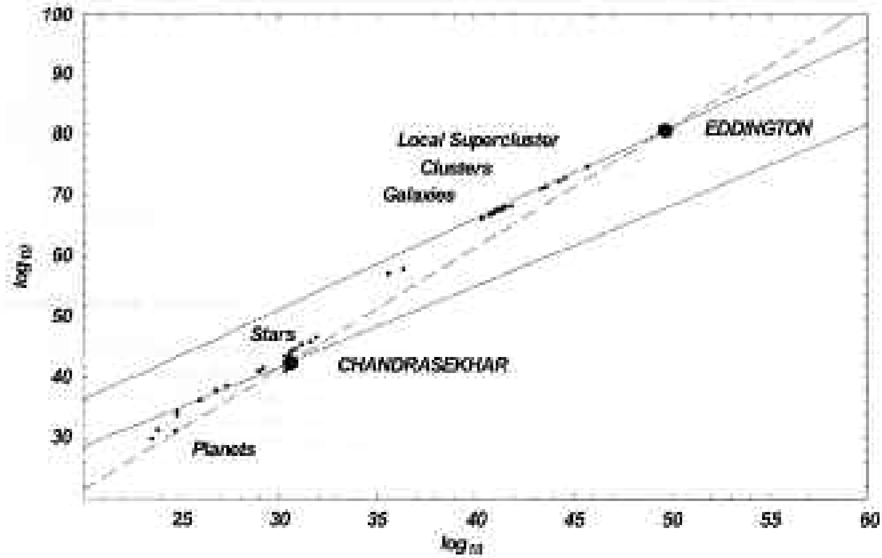


Fig. 3. Joining observation and theory. The angular momentum of planets and stars is well described by the ball Regge trajectory. Galaxies and their clusters are very close to the disk trajectory. Two interception points of the Kerr spin (dashed line) are named Chandrasekhar and Eddington points and their coordinates have fundamental meaning.

This answer is correct, but seems incomplete. We must add that the 'stars are as they are' because also the combination

$$\hbar\left(\frac{\hbar c}{G m_p^2}\right)^2$$

provides a correct measure of stellar spins.

Remark about 'dark matter'. The possible existence of dark matter within the galactic halo at first was proposed by Fritz Zwiky some 40 years earlier in his studies of behavior of galaxies in clusters. This deduction was made from the premise of the stability of clusters and thus of the applicability of the virial theorem. An alternative explanation was proposed by Soviet Armenian astronomer Victor Ambartsumian [13] who considered the possibilities for the existence of a non-stability and (nongravitational) source of total positive energy in clusters. The same reasoning could be applied to the problem of flat rotation curves in spiral galaxies. We need no dark matter to explain the observed flatness of rotation curves if rotation is non-stationary and some unknown source continuously creates angular momentum within a galaxy. It is possible to attribute the increase of angular momentum to the activity of the nuclei of galaxies. The supermassive remnant with a Regge-like spin, located at the center of a galaxy could be a primary reason for the rotational activity of galaxies. The challenge of 'missing mass' and 'missing angular momentum' is serious, but not fatal. As noted e.g. in [14]: 'The need for missing mass disappear if one admits that galaxies and galactic clusters might not be in (dynamic) equilibrium'.

4. Concluding Remarks

This talk serves to demonstrate in what way power scaling laws could be equally useful in the description of physics at very small and very big distances. The success of these applications is witnessed by the unity and simplicity of Nature in the range from elementary particles up to clusters of galaxies and the Universe itself.

REFERENCES

1. Muradian, R., Urintsev, A., *Diana*: *a Mathematica Code for Making Dimensional Analysis*, preprint JINR E2-94-110, 1994.
2. Matveev, V., Muradian, R., Tavkhelidze, A., *Nuovo Chimento Letters*, 7, 719 (1973).
3. Lepage, G., Brodsky, S., *Phys. Rev*., D22, 2157 (1980).
4. Bjorken, J., *Phys. Rev.,* 179, 1547 (1969).
5. Matveev, V., Muradian, R., Tavkhelidze, A., preprint JINR, P2-4578, 1969.
6. Lee, T.D., *High Energy Electromagnetic and Weak Interaction Processes*, preprint Columbia University CRISP, 71-57; *Scaling Properties and the Bound-State Model of Physical Baryons*, preprint CERN, 73-15, Geneva, 1973.
7. White, G., *et al.*, *Phys. Rev.,* D49, 58 (1994).
8. Muradian, R., *Astrofizika,* 11, 237 (1975).
9. Muradian, R., *Astrofizika,*13, 63 (1977).
10. Muradian, R., *Astrofizika,* 11, 439 (1978).
11. Muradian, R., *Astrophysics Space Science*, 69, 339 (1980).
12. Muradian, R., *Regge in the Sky: Origin of the Cosmic Rotation,* preprint ICTP IC/94/143, 1994.
13. *Non-stable Phenomena in Galaxies*, proceedings of IAU Symposium n. 29, Byurakan, May 4-12, 1966, The Publishing House of the Academy of Sciences of Armenian SSR, Yerevan, 1968.
14. Corliss, W.R., *Stars, Galaxies, Cosmos,* The Sourcebook Project, Glen Arm, MD, 1987.

PATHS OF DISCOVERY: PERSONAL EXPERIENCES IN A SOCIAL SCIENCE

BERNARDO M. COLOMBO

The foundation of Demography may be traced to a booklet of 'Natural and Political Observations' made on the Bills of Mortality, then in use, that John Graunt presented to the Royal Society in 1662.[1] The death of a person is a common topic in a social conversation. Graunt, observing that, beyond that, little use was made of the information recorded in the Bills, considered the possibility of extending it in other directions and for other purposes. The 'death' event became for him simply an abstract object with certain characteristics: location, time, cause of the event, sex and – roughly – age of the deceased person... He went on, making classifications of those objects according to the specific expression of those characteristics in each observed unit collected over a number of years in the city of London and in the country town of Romsey, New Hampshire. In each science, the purposeful classification of observations is a fundamental step. As to the collection of data, in the area of population it goes back to ancient times. As we know, the birth of Jesus is linked to an operation of that kind. The classification of individuals is also old. A Psalm reminds us that 'The days of our life are seventy years – or eighty if we have the strength' (NRSV, *Ps* 90:10). A text that implies classification by age at death.

Graunt took a further step, looking for configurations and regularity in time of observed phenomena, but also putting in relation different classifications for the purpose of enriching knowledge. May I quote one example.[2] He saw that, year after year, the numbers of abortions and still-

[1] Graunt J., *Natural and Political Observations Mentioned in a following Index, and made upon the Bills of Mortality,* 1662, London.

[2] Glass D., 'John Graunt and his Natural and Political Observations', *A Discussion on Demography*, Proc. of the Royal Society, B., vol. 159, 1963, 2-32.

births were roughly constant. The number of christenings – taken as the number of births – after a normal period, started and continued declining over several years. Taking the figures of 1631 – a normal year without problems – he calculated the ratio of abortions to christenings. Assuming that this ratio remained constant in time, he made an estimate of the 'true' number of christenings in 1659, a recent year in which he thought that 'the neglect in the *Accompt* of christenings' continued. To check this estimate, he followed a similar procedure using the number of women who died in childbirth. The techniques of analysis of the data he used were plain exercises of arithmetic, but as a path to discovery they were 'outstanding innovations'.

To advance one more step, I take a modern example. It concerns the measurement of the level of fecundability, taken as the probability that a married woman has a conception during a month, making abstraction from any practice meant to limit procreation. This concept, suggested in 1923 by Corrado Gini, opened a new chapter in Demography.[3]

According to Gini:

p = is the probability that a woman conceives in a month, probability supposed constant in time and the same for all women

n_x = is the number of women who did not yet have the first birth up to the x^{th} month since marriage.

Then $p \cdot n_x$ = number of first born by these women during the month x. The number of first births during the month $x + 1$ – disregarding mortality – will be $(1 - p) p \cdot n_x$ and during the month $x + 2$ will be $(1 - p)^2 p \cdot n_x$. That is, the number of first births in successive months forms a geometrical progression of ratio $(1 - p)$. Month by month you easily get a rough estimate of p. A very simple operation, concerning which I highlight two aspects.

First, it is possible because each single unit is cross-classified by parity and distance from marriage. Cross-classification of covariates is an essential tool for understanding phenomena influenced by the concomitant impact of several factors. In the field of population studies, it is a demanding task. Take, for instance, a Yearbook of vital statistics based on documentation taken usually from administrative records. There are many variates of interest and a huge number of possible cross-classifica-

[3] Gini C., "Prime ricerche sulla 'fecondabilità' della donna", *Atti del R.*, Istituto Veneto di Scienze, Lettere ed Arti, Anno Accademico 1923-24, vol. 58, Parte 2, Venezia, Ferrari, 1924, 315-344.

tions with a variety of details. For the book, you must make a choice among them. This requires ideas about what might be relevant for scientific investigation. That demands a great amount of culture and imagination about the course of science. In the Yearbook you are providing building blocks for future research. According to the choices you make, you open or close doors. From personal experience, I know how challenging these choices are.

Secondly, besides the availability of adequate data, you need intuition on how to use them to shed light on unexplained or unknown realities. But how do such ideas come?

May I quote a personal experience. As a young research assistant in the University of Venice, I was confronted with the problem of understanding the reasons for the recovery of the birth rate in several countries during the Second World War: countries directly involved in the operations, or neutral but near the regions where battles were going on. An English colleague had seen in that a making up of births delayed during the great economic crisis, without a real change in the final family size.[4] In order to test this hypothesis, in each instance a series of specific rates standardized with respect to both parity and distance from marriage had to be computed. Detailed tables – with appropriate cross-classifications – were usually published each year, providing the figures for the numerators of such rates. The data for the denominators were missing everywhere, even in census years. I was at this point when one evening I went to the movie theatre to relax. Walking back home, with that problem in mind, I suddenly had an idea about how the data available for the numerators could, through appropriate elaborations, be used to create the figures for the denominators of those rates. It worked. I thus had the possibility of showing that fertility in those countries really increased, without any making up.[5]

I also tried to give an explanation of that phenomenon in a few pages of considerations which I supposed plausible. When the teacher who was my guidance finished reading the manuscript, he asked me: 'Can you prove that?'. I had to answer: 'No'. Those pages were omitted in the published paper. That lesson made me very cautious in evaluating conclusions

[4] Hajnal J., *The Analysis of Birth Statistics in the light of the recent International Recovery of the Birth-Rate,* Population Studies, 1, 2, 1947, 137-164.

[5] Colombo B., *La recente inversione nella tendenza della natalità,* CEDAM, Padova, 1951, VIII+183.

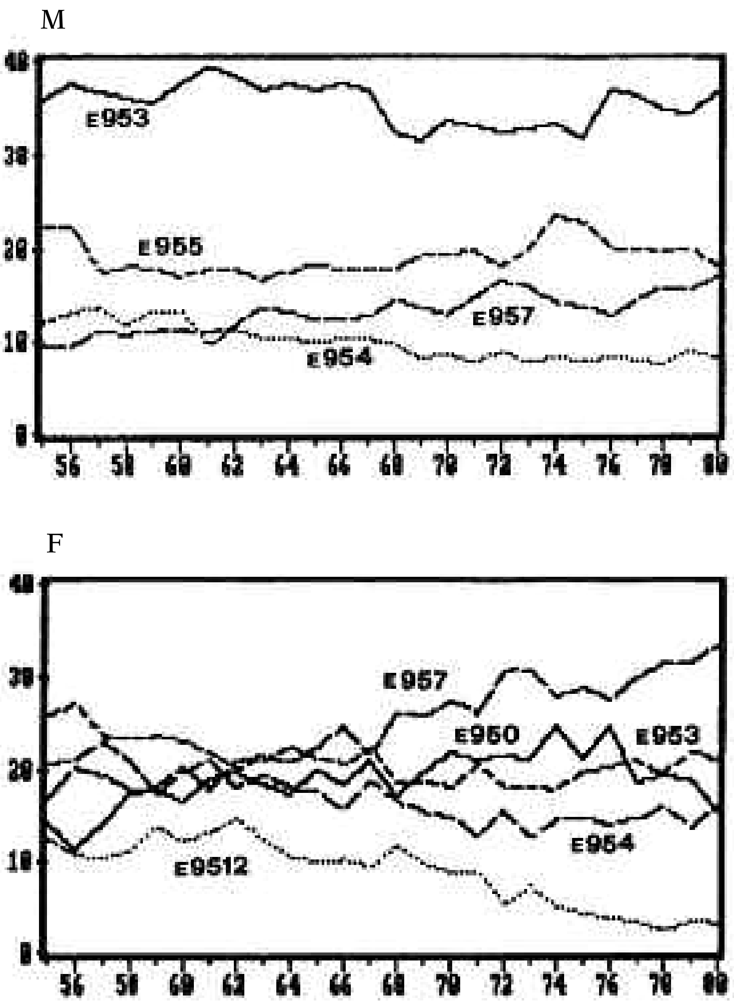


Fig. 1. Percentage of suicides of various categories on the total number of suicides for the given sex: Italy, 1955-1980.
Note: Suicides and self inflicted injury by
E950= poisoning by solid or liquid substances
E951= poisoning by any gas in domestic use
E952= poisoning by other gases
E953= hanging, strangulation and suffocation
E954= submersion (drowning)
E955= firearms and explosives
E957= jumping from high places
E9512= E951+E952

reached in similar areas. The risk of spurious correlations, ecological fallacies, explanations of the 'post hoc ergo propter hoc' type, and so on, now immediately raise my critical attention.

Demographic analysis can offer clear illustrations of ongoing phenomena. Take, for instance, the baby boom of the postwar period, a development common to Western Countries. Its peak was reached in 1960 in North America and in 1964 at the same time in several countries of the European region. But when you ask: 'Why is that?', demographic analysis alone is unable to give answers.

Or take the deaths due to suicides recorded in a recent quarter of a century in Italy – from 1955 through 1980.[6] Charting the figures of the two sexes according to the different forms of self-murder, one can see that for the males the lines corresponding to the different manners run parallel over the years, while for the females they change levels and even cross (see Fig. 1). Maybe the collaboration of different disciplines can shed some light on the relevant factors having an impact on these social trends, but within a limiting condition. In this field, experiments are denied. In the interpretation of social events you can hardly give unequivocal proof, as you are unable to falsify it.

Population dynamics is at the crossroads of physiology and behaviour and raises problems of interpretation on both sides. Take, for instance, the sex of newborns, with the slightly higher frequency of males, the most well-known observation made by John Graunt. I studied in depth the question of the systematic differences occurring in the comparison of chosen selected groups of births classified by sex.[7] I found confirmed, for example, in the Second World War, what had been observed in the first one: an increased masculinity ratio. I will not elaborate on the hypotheses made by various authors about factors thought to be behind the results. Surely, to confirm or disprove any of them, a third world war would be too costly an experiment. Similar intriguing evidence defies interpretation in several other instances. Many hypotheses have been advanced to provide explanations, hypotheses that often died with their authors, so that we can repeat what Sir Francis Bacon happened to state: 'What was a question once, is a question still'.

[6] Colombo B., Fortuna I., 'La regolarità degli eventi rari in Demografia', *Atti e Memorie dell'Accademia Patavina di Scienze, Lettere ed Arti*, Parte III, 1992, 65-83.

[7] Colombo B., 'On the sex ratio in man', *Population Studies: Animal Ecology and Demography*, Cold Spring Harbor Symposia on Quantitative Biology, 22, 1957, 193-202.

Along the same line, I take another example from personal experience. Making use of the very good and sizeable Italian vital statistics, I found that the probability of multiple births was increasing regularly and independently of both age of the mother and parity. At higher ages and parities it reached a level of more than four times the initial lower one.[8] A colleague showed that this trend depended on dyzigotic twins.[9] I contacted a geneticist, asking for an explanation: he had none.

Certainly demography must not be taken as a technique specialized in raising problems. It offers in fact, powerful tools to understand what is happening today and to foresee the scenarios of the future. Through Lotka's stable population theory and the model life tables of Coale and Demeny, for instance, it gives the possibility of deriving from small and uncertain pieces of information a complete and coherent picture of the general structure and dynamics of a population. This path is useful for understanding the situation of countries with poor or missing data.

The advent of the computer has drastically enriched the instruments of analysis of the discipline allowing to go to the roots of what we see happening at the population level. Births, deaths, marriages, migrations depend on individual decisions or conditions, in the context of a specific milieu. Now, using all the information provided by single units, we can try to derive from them, through appropriate models and procedures, how and why things follow the trend we see in the general population.

Intuition and imagination are here involved at various steps: in the clear expression of the question to be answered; in the identification of the variables – and their possible interactions – to be taken into consideration; in the definition of a model which – through a regression equation – links them to the empirical observations that need to be explained. A model may carry out several functions: it gives a synthetic description of ascertained evidence; it clarifies the links between variables and the causal relations with the observed data; it allows us to make predictions. Within the limit of stability of the characteristics of single units and of the involved links with specified covariates, it makes it possibile to foresee future developments. But it also allows to make, through extrapolation,

[8] Colombo B., *Appunti di Statistica per ostetrici e ginecologi,* Clinica Ostetrico-Ginecologica, Università di Padova, Anno Accademico 1960-61.

[9] Rigatti Luchini S., 'La frequenza dei parti monozigotici e dizigotici, in relazione all'età della madre al parto e all'ordine di nascita, in Italia', *Atti della XXIX Riunione scientifica Società Italiana di Statistica*, vol. 2, 2, Bologna, 1978, 73-83.

extensions to peculiar sets of population: for instance to those provided with extreme features of characteristics considered relevant. This last exercise needs to be done very carefully. What you are doing, in fact, are elaborations on virtual populations. If you want to be comfortable with the results, you need the support of some evidence.

ON A DISCOVERY ABOUT GÖDEL'S INCOMPLETENESS THEOREM

STANLEY L. JAKI

Ever since René Taton came out in 1955 with his book *Causalités et accidents de la découverte scientifique*,[1] which quickly appeared in English as *Reason and Chance in Scientific Discovery*,[2] much interest has been shown in the nature and circumstances of scientific discoveries. In that now vast literature it has often been noted, for instance, that a discoverer does not always perceive the full meaning of what he had discovered. On occasion that literature contains a discussion of the fact that discoveries can also be resisted, which is, of course, one of the worst things that can happen in science. The first egregious case was provided by Galileo in reference to Kepler's laws. Then there came the notoriously long resistance to the wave theory of light. To discover such things belongs to the historian of science.

The history of physics during the twentieth century began with Planck's resistance for more than ten years to the notion that he had discovered the quantum of energy, or rather the fact that atoms absorb and emit energy in a discontinuous way.[3] Einstein did not wish to consider for years the growing evidence, theoretical and observational, of the expansion of the universe. About the recently deceased Thomas Gold, one of the first proponents, with F. Hoyle and H. Bondi, of the steady state theory, it was prominently recalled in the *Economist* (July 3, 2004, p. 73) that he resisted to the

[1] With the subtitle, *Illustrations de quelques étapes caractéristiques de l'évolution des sciences* (Paris: Masson).

[2] New York: Philosophical Library, 1957; then issued in paperback (New York: Science Editions, 1962).

[3] Many details about that resistance can be found in T.S. Kuhn's *Black-Body Theory and the Quantum Discontinuity 1894-1912* (Oxford: Clarendon Press, 1978). See also ch. 1 in my book, *Numbers Decide and Other Essays* (Pinckney, Mich.: Real View Books, 2003).

end the devastating blow dealt to that theory by the findings of COBE (Cosmic Background Explorer). It is another matter that the notion, according to which hydrogen atoms are popping up everywhere in cosmic spaces out of nothing, should not have been resisted as something that cannot be dealt with by the methods of physics. This is so because the 'nothing' and the process of coming into being out of nothing cannot be measured.

In all such cases of resistance various factors – psychological, sociological, and ideological – play a role. The first two factors have been extensively investigated, and at times to the point bordering on the tedious and the ridiculous. Some created the impression that in order to make a significant scientific discovery one has to be twenty-five or younger, or come from a working class family, or have a contempt for social conventions. Practically no attention has been given to the ideological factor, possibly because ideology is akin to religion and in this age of ours religion is considered to be a strictly private matter, if it is considered at all.

Yet ideology seems to have played an important role in the resistance by prominent physicists to perhaps the greatest discovery in the history of mathematical logic, or Kurt Gödel's formulation, in November 1930, of the theorem that any non-trivial set of arithmetic propositions has a built-in incompleteness. The incompleteness consists in the fact no such set can have its proof of consistency within itself. The bearing of that incompleteness on physical theory, which has to be heavily mathematical, should seem obvious.

Physicists had all the less reason to ignore Gödel's discovery, because it was published in *Monatshefte für Mathematik und Physik*,[4] a leading monthly which carried as many articles about the latest in mathematics as in physics. Yet, thirty-five years went by before attention was called to the bearing of Gödel's theorem on the formulation of a comprehensive physical theory. As long as Godel's theorem remains valid, the formulation of a final, or necessarily true physical theory should seem impossible. This is so because such a theory, whether it is called Unified Field Theory, or the Theory of Everything (TOE), also called M theory,[5] cannot have its proof of

[4] 'Ueber formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I', Volume 38, 1931, pp. 173-198. or 'Formally undecidable propositions of *Principia Mathematica* and related Systems I'. The German original and its English translation are available on facing pages in *Kurt Godel: Collected Works. Volume I. Publications 1929-1936*, ed. S. Feferman (Oxford: Clarendon Press, 1986), pp. 144-95.

[5] The M stands for a variety of words, such as Master, Majestic, Mother, Magic, Mystery, Matrix. See B. Greene, *The Fabric of the Cosmos: Space, Time and the Future of Reality* (New York: Alfred A. Knopf, 2004), p. 379.

consistency within itself. Yet a necessarily true theory cannot lack the quality of inner consistency in its mathematical part.

When this line of reasoning appeared in my book *The Relevance of Physics*, published by the University of Chicago Press in 1966,[6] the TOE theory and the M theory were still in the future, but much had been spoken of a Unified Field Theory, mainly a brainchild of Einstein. The context of my reasoning, or discovery, relates to the first Part of *The Relevance of Physics*, where in three chapters I dealt with the three main types of physics that prevailed in Western intellectual history. Those three types are the organismic, the mechanistic and the mathematical. The first, in which the world was taken for a living organism, prevailed from Aristotle on until Copernicus, Galileo, and Newton. The second, or mechanistic physics, in which interactions among particles of matter were viewed as interaction among bits of machinery, dominated the thinking of physicists until the beginning of the 20th century.

From the early 20th century on the idea has gained ground that the physical world was ultimately a construct in numbers. Consequently, the mathematical perfection of physical theory has been increasingly taken for an index of the perfection of the physical theory itself. Indeed by the 1960s many physicists voiced the hope that a final form of physics, or a final formulation of fundamental particles, would soon be on hand. Many such hopeful expressions are quoted in chapter 3 of *The Relevance*, a chapter which has for its title, 'The World as a Construct in Numbers'. At the end of that chapter I argued that because of Gödel's theorem such hopes were without foundation.[7]

In none of the hundred or so reviews of *The Relevance* was attention called to that connection between Gödel's theorem and a perfect or necessarily true physical theory. And much the same is true of the reviews of still other four books of mine in which I set forth the same line of argument, prior to 2002. Among those books were my Gifford Lectures, *The Road of Science and the Ways to God*,[8] given at the University of Edinburgh in 1975 and 1976. After that I set forth the arguments in *Cosmos and Creator* (1980),[9] then in my Fremantle Lectures, *God and the Cosmologists*, given at

[6] Two more editions followed, the last by Scottish Academic Press (Edinburgh) 1992.

[7] *Ibid.*, pp. 127-129.

[8] Chicago: University of Chicago Press, 1978. See pp. 253, 427, 453, and 456. This work was also published simultaneously by Scottish Academic Press, and brought out in paparback by both Presses in 1981.

[9] Edinburgh: Scottish Academic Press, 1980. See pp. 49-51.

Balliol College in 1989,[10] and finally in my Forwood Lectures, *Is there a Universe?*, given at the University of Liverpool in 1992.[11] I should also mention a paper I read at the Nobel Conference at Gustavus Adolphus College in Minnesota in 1976.[12] The five other members of the panel were Fred Hoyle, Steven Weinberg, Hilary Putnam, Victor Weisskopf, and Murray Gell-Mann. What happened there when I referred to Gödel's theorem I told in detail in a paper just published.[13] What I said in all these publications about the connection of Gödel's theorem and a perfect physical theory has been ignored with only one exception. In his book *Impossibility*, J.D. Barrow quoted some lines from my *The Relevance of Physics* and *Cosmos and Creator* and in both cases he misconstrued what I said.[14]

One may perhaps say that the persistent neglect of what I said over thirty years about Gödel's theorem and physics was not valid. But then one has to say the same about a paper which Prof. Stephen Hawking read at the centenary celebration of Dirac's birthday, held in the Centre of Mathematical Sciences at Cambridge University on July 23, 2002. The very title of his paper, 'Gödel and the End of Physics',[15] should have created enormous attention well beyond the world of physicists, and should have brought attention to Gödel's discovery made seventy-two years earlier. Nothing of

[10] Edinburgh: Scottish Academic Press, 1989. See pp. 84-110.

[11] Published by Liverpool University Press in 1995. See pp. 101 and 107.

[12] 'The Chaos of Scientific Cosmology', in D. Huff and O. Prewett (eds.), *The Nature of the Physical Universe: 1976 Nobel Conference* (New York: John Wiley, 1978), pp. 83-112. Published also in Italian translation, 'Il caos della cosmologia scientifica', in *Natura dell'universo fisico* (Torino: P. Boringhieri, 1981), pp. 88-114.

[13] See note 15 below. Following the presentation at the same conference by Murray Gell-Mann, who promised a two-thousand strong audience that within three months, or certainly within three years, he would come up with a final theory of fundamental particles, I reminded him of Gödel's theorem and of the consequent futility of his project. It was then that he first heard of Gödel. But two months later he gave a paper at the University of Chicago and stated that because of Gödel's theorem that final theory cannot be achieved. Only he failed to refer to the incident at the Nobel Conference.

[14] With the subtitle, *The Limits of Science and the Science of Limits* (New York: Oxford University Press). In neither case did I say that, as Barrow would have it, I see Gödel's theorem to be 'a fundamental barrier to understanding the Universe'. It is such a barrier only for those who want a 'final' understanding of the universe in terms of mathematical physics and cosmology.

[15] Made available on the Internet via 'ogg orbis', and discussed in my essay, 'A Late Awakening to Gödel in Physics', *Sensus communis* 5 (2004) 2-3, pp. 153-162, available also on my website, www.sljaki.com. The article just appeared in Hungarian translation in *Fizikai Szemle* (Budapest). An Italian translation is forthcoming.

this happened. Hawking's paper created no significant echo. Yet that paper of his was far more fundamental for physics than the paper he presented on July 21, 2004, at the 17th International Conference on General Relativity and Gravitation. This paper, in which Prof. Hawking reversed his long-standing opposition, with an eye on his theory of black holes, to the principle of time reversal, was announced in headlines all over the world. The headline in *The New York Times* was 'Those Fearsome Black Holes? Dr Hawking Says Never Mind'.[16]

But this paper is not about the history of debates about black holes but about reasons of a widespread resistance to what I have kept saying about Gödel's theorem and a perfect, or necessarily final physical theory. The basic resistance has distinctly ideological aspects, and this is all too obvious in Hawking's case. He had repeatedly stated his atheism and indeed boasted of it. Now for an atheist or materialist there can be only two final entities: either his own mind or the material universe. In the case of physicists (or cosmologists) who dream of a final theory, the final entity is usually their own mind. Let me begin with a remark Prof. Hawking made in 1976, when he retorted Einstein's words, 'God does not play dice', with the remark that 'God not only plays dice with the universe, but sometimes throws them where they cannot be seen'. Only some amateurs in theology would be impressed by such remarks, made either by Einstein or by Hawking. Both were atheists in their own ways, which in their cases too implies some consequences for their patterns of thinking.

Theology, or rather a hint about Hawking's pseudo-theological motivations, again surfaced when a brief discussion appeared in *The New Scientist* of Hawking's paper, 'Gödel and the End of Physics', though only almost a year after its presentation. On the cover of the April 5, 2003, issue of *The New Scientist,* one could read the double caption, 'The Mind of God' and 'Hawking's Epiphany', of which at least the first was clearly theological. The reason for bringing in theology, and in a headline at that, related to the claim Prof. Hawking made in 1988 in his book *A Brief History of Time*, that a boundary-free physical theory makes the Creator unnecessary.[17] The claim spread far and wide because *A Brief History of Time* quickly became a runaway bestseller of perhaps all times. Within four years it sold in five million copies and is still selling. My concern here is not about that claim's illogicalities both from the viewpoint of physics and theology, which I dis-

[16] July 22, 2004, pp. A1 and A3. The report was written by D. Overbye.
[17] *A Brief History of Time* (Toronto: Bantam Books, 1988), pp. 173-74.

cussed elsewhere.[18] I am concerned here with the theological roots of a scientific resistance to the bearing of Gödel's theorem for physics, as an illustration of some broader aspects of scientific discoveries. And this resistance was clear in the article of *The New Scientist*. There one would look in vain for a reference to the fact that it was with an eye on Gödel's theorem that Prof. Hawking reversed his claim about a boundary-free physical theory. One has to go well beyond *The New Scientist* and even a *Brief History of Time* to see the depths of Prof. Hawking's resistance to Gödel's theorem. When that book of his was published in 1988, two years had gone by since the publication of *Kurt Gödel. Collected Works. Volume I. Publications 1929-1936*.[19] There in the Introduction one reads that Prof Hawking is one of the authors of introductions to various papers of Gödel. Volume I contains no such introduction by Prof. Hawking, who most likely was asked to write an introduction to Gödel's paper on rotational cosmology.[20] No second volume of that *Collected Works* has so far appeared. But since according to the same main Introduction to Volume I, that Volume had long been in preparation, one may safely assume that as early as 1980 Prof. Hawking's attention was explicitly called to Gödel's work.

At any rate it would be well nigh impossible to assume that already during his student years in Cambridge, that is, the 1960s, Hawking would have remained unaware of Gödel's paper of 1931 which contains the incompleteness theorem, and the less so as the theorem was widely discussed in England in the 1950s and 1960s in connection with debates on artificial intelligence.[21] Moreover, it was in Great Britain that the best English translation of Gödel's paper was published in 1962, with an introduction by the Cambridge philosopher of science, R.B. Braithwaite.[22]

[18] 'Evicting the Creator', (1988); reprinted also in my *The Only Chaos and Other Essays* (Lanham, Md.: University Press of America, 1990), pp. 152-161.

[19] Under the editorship of Solomon Feferman (New York: Oxford University Press; Oxford: Clarendon Press, 1986).

[20] It appeared in two instalments. The first, 'An example of a new type of cosmological solutions of Einstein's field equations of gravitation', *Reviews of Modern Physics* 21 (1949), pp. 447-50; the second, 'Rotating universes in general relativity theory', in *Proceedings of the International Congress of Mathematicians, Cambridge, Massachusetts, USA August 30-September 6, 1950* (Providence, R.I.: American Mathematical Society, 1952), vol. 1, pp. 175-181.

[21] See my *Brain, Mind and Computers* (1969), 3d enlarged edition (Washington: Regnery Gateway, 1989), pp. 214-16.

[22] Kurt Gödel, *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*, trans. B. Meltzer, with an Introduction by R.B. Braithwaite (Edinburgh: Oliver & Boyd, 1962).

But there is one more fact, of which I learned only in June 2004. It was then that out of the blue I was contacted by Mr John Beaumont, formerly the legal counsel to Leeds University. Though not a physicist, Mr Beaumont has for years followed closely the developments in physics. He read *A Brief History of Time* shortly after its publication. Two years later he bought a copy of my book, *God and the Cosmologists*, or the text of my Fremantle Lectures, given in Oxford. In that book the entire chapter 4, with the title 'Gödel's Shadow',[23] deals with the bearing of Gödel's theorem on physics. My book alerted Mr Beaumont to the significance of Gödel's theorem for physics and he informed Prof. Hawking about it. To that communication Mr Beaumont received no reply.[24]

So much about my claim, if it had to be made at all, that Prof. Hawking had for almost two decades been aware of Gödel's theorem, before he took it up in June 2002, in reference to physics. Worse, when he did so, he made the impression that he was the first to do so. At any rate, he made the erroneous claim that Gödel's theorem means the end of physics. It means exactly the opposite. A physicist may hit upon a theory which at a given moment could cope with all known problems and phenomena in physics. But he cannot reasonably expect that in the future nothing will be observed that would require a radical overhaul of the theory. And because of Gödel's theorem, the physicist in question cannot claim that the apparent mathematical perfection of the theory forecloses that possibility. In other words, precisely because of Gödel's theorem there will be no end to physics.

The purpose of this paper is not to vindicate my priority about the discovery of the bearing of Gödel's theorem to physics. Actually, sometime in the early 1970s I saw a book on physics, published a few years earlier than my *Relevance,* whose author stated briefly that because of Gödel's theorem it was not possible to formulate a final physical theory. Unfortunately, before I had time to write down the title of that book, it disappeared from my eyes. It may indeed be that there were other such books as well, a possibility which some historians of physics may find worth investigating.

The purpose of this paper was to probe into some ideological reasons about a strange resistance on the part of leading physicists to the connection between Gödel's theorem and the possibility of formulating a necessarily true physical theory. Given the heavily agnostic temper of thinking

[23] See note 10 above.

[24] This I know from an email, sent to me by Mr Beaumont in early June 2004, and entirely at his own initiative.

among leading physicists, I am not surprised that what I said failed to prompt a proper reaction on their part. But it is even more significant that they have equally ignored Hawking's paper of 2002. In fact they ignored even the report of that paper which appeared in the April 5, 2003, issue of *The New Scientist*. Why is it, one may ask, that whereas Prof. Hawking's very recent rejection of his theory of black holes makes news all over the world and appears on the front page of leading newspapers, his paper 'Gödel and the End of Physics', prompts no response on the part of physicists who work very much on a final theory?

Surely, it is not possible to assume that Prof. Brian Greene of Columbia University has not yet heard of Gödel's theorem, or that he has not heard of Hawking's much belated awakening to that theorem. When *The New Scientist* reported about it, Prof. Greene, best known for his work on superstring theory, was just finishing his book, *The Fabric of the Cosmos*.[25] The book is full of hopeful expressions that the string theory would be the final word in physics and an explanation of everything not only in physics but of everything beyond. Prof. Greene's failure to refer there to Gödel's theorem is all the more telling because he discussed over two pages Gödel's rotational model of the universe.[26] I cannot help thinking that Prof. Greene simply refuses to consider Gödel's theorem. The refusal here, too, is distinctly ideological in its nature, which becomes clear to anyone who pays attention to Prof. Greene's repeated endorsement of some form of materialism throughout his book. Again, there is much more than meets the eye in Prof. Greene's earlier book, *The Elegant Universe*, which, to make the irony complete, has the subtitle: *Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*.[27]

Once more elegance was purchased at the price of substance. A poor bargain, beneath which there must lie strong motivations. Prof. Greene seems ignoring Gödel's theorem for a reason that should be obvious. As in his *Elegant Universe*, in *The Fabric of the Cosmos*, too, Prof. Greene repeatedly makes it known that for him the universe is the ultimate entity. For anyone with such a belief Gödel's theorem should be a thing most distasteful to consider. I have a certain measure of sympathy for that distaste. It is a little known secret about the history of philosophy, that no philosopher of

[25] With the subtitle, *Space, Time and the Texture of Reality* (New York: Alfred A. Knopf, 2004).

[26] *Ibid*., pp. 259-60.

[27] New York: W.W. Norton, 1999, xv + 448 pp. A second edition followed in 2003.

any stature has ever changed his basic convictions. Conversions are rare in philosophy and even rarer in matters theological. While the number of converts may be large, the number of converts who had considerable theological views is rare indeed. This holds true also of those, scientists or not, who hold strongly anti-theological views. And in this age of science when anything wrapped in science readily sells, nothing is so tempting than to wrap one's wares in science. It should therefore be difficult for some scientists, who construct a scientific cover for their antitheological views, to recognize their mistaken tactic.

My dispute is not with non-theistic or atheistic ideologies. Ideology, in a broad sense, is inevitable as long as one thinks. Ideology is always present at least in the form which a century and half ago became referred to as one's antecedent and often tacit assumptions. It is very important to probe into those assumptions if one is to see clearly the full picture about one's philosophical theory. Had the recently deceased Jacques Derrida done this he would have made a useful contribution to philosophy. Instead he cavorted in what has become known as 'deconstructionism', where for all practical purposes confusion is proposed as the means of enlightenment.

Tacit assumptions rule the final outcome of any debate. Therefore it is best to be clear and honest about them. A few years ago I suggested in this Academy that at meetings such as this, all participants should wear a necktie appropriate to their tacit assumptions. I should have, of course, known that such a suggestion would never be taken up. Yet the very constitution of this Academy calls, if not for the wearing of specific neckties, at least for an occasional airing of basic assumptions. This is so because the constitution of this Pontifical Academy calls for discussions on the relevance of this or that scientific finding on this or that doctrine or dogma of the Catholic Church. In other words, it is proper to consider here the relation of science and faith within the framework of at least the Plenary Meetings of the Academy. I am somewhat reluctant to mention this, because during the past fourteen years I have been coming to these meetings some contrary views have been voiced by some academicians in this aula.

About the relation of science and religion much nonsense has been offered ever since the rise of science four hundred years ago. One such nonsense is that they are in harmony or even in a sacred wedlock, to recall a famous remark from Newton's time. They are neither in harmony, and much less find themselves in such a wedlock. They cannot be in harmony, because they are about two different sets of propositions. Kepler in his

Harmonices mundi could dream about the music of planetary orbits, but that book would not have helped composers. Conversely the magnificent tonalities in Haydn's Oratorio 'The Creation' contained no clues for science. It is of little significance to claim that the first law of motion was first formulated in a distinctly theological context at the Sorbonne around 1330. Although this was enormously important for the future of physics, it would be a great exaggeration to say simply that modern science was born in a Christian context.[28] Modern exact physical science, as we find it first in Newton's *Principia*, has been cultivated with no reference to Christian theology whatsoever, insofar as that science was strictly physics and not some broad thinking about physics and the physical world. Equally misleading in the cliché about science and religion is that they are in fundamental conflict, in a warfare indeed, to recall a shibboleth very popular in the second part of the nineteenth century.[29]

There is in my mind only one serious objection that science can make to a religion which is much more than a worshipping of great nature. I do not think that the objection is serious, but in this age of science everything wrapped in science calls for a special consideration. I do not think that a final theory, even if necessarily true, would render the Creator unnecessary. That final theory still would have to deal with the incredibly high degrees of specificities everywhere in the material universe. Specificities are all cases of finiteness, that is, of restriction. They all provoke Leibniz's question: 'Why such and not something else?'.[30] Infinite perfection does not have to be explained. It is finite things that call for explanations and indeed they alone prompt calls for explanations. Those appreciative of the profundity and weight of Leibniz's question will alone cherish Chesterton's remark that the sight of a telephone pole is enough to prove the existence of a Creator.

But we live not in a philosophical but in a scientific age. Therefore for the benefit of those who can take seriously only science and practically nothing else, it is useful to point out that the idea of a necessarily final

[28] For such a restriction, see my *The Origin of Science and the Science of its Origin* (Edinburgh: Scottish Academic Press, 1978) and my *Means to Message: A Treatise on Truth* (1999), and especially my *Questions on Science and Religion* (Port Huron, Mich.: Real View Books, 2004).

[29] See on this my introduction to the re-edition of K.A. Kneller's *Christianity and the Leaders of Modern Science* (1911) by Real View Books (1995), pp. xiii-xiv.

[30] On that question of Leibniz in his essay on the ultimate origination of things (1697), which remained in manuscript until 1840, see my *Cosmos and Creator*, pp. 90-92.

theory is beset with an enormous defect, the defect being its mathematical incompleteness in terms of Gödel's theorem. Were that defect not there, minds attentive only to science might argue that the Theory of Everything (TOE) deprives the universe of its contingency. In that case the kind of religion, which begins with belief in the existence of a Father Almighty, would lose its credibility, though only within that very narrow perspective coated with science.

For those who think only within that narrow perspective and in addition entertain antitheological antecedent assumptions, the specter of Gödel's theorem may not be pleasant to contemplate. They have various ways for protecting themselves from its specter. One is, for instance, to write books with the title, *The Mind of God*,[31] and earn handsome royalties. They react to a mere whiff of real theology as King Lear looked at an ominous prospect and cried out: 'Oh, that way madness lies; let me shun that'. There is, to paraphrase Shakespeare, a madness in a method which demands resistance to the obvious. Another dubious method is to follow the example of the ostrich, which, according to the fable, buries its head when faced with an approaching enemy. It seems to me that the long-standing resolve to ignore Gödel's theorem shows something of the tactic which all real specimens of the avian kingdom have wisely refused to adopt.

Gödel's theorem does not mean the end of physics. On the contrary it assures physicists that they can work forever for a final theory, though with one important proviso. Even if they were to hit upon that final theory, they would not know with certainty that it is final. This lack of finality in physics has in Gödel's theorem a stronger proof than the very likely expectation that, as time goes on, experiment and observations would turn up data that would demand the complete overhaul of well-established theories, just as this happened a hundred years ago. This seems to be the argument in Prof. Weinberg's book *Dreams of a Final Theory* in which he pokes fun on dreams about them, while he keeps silent on Gödel's theorem,[32] although it is the only solid argument against final theories.

On essentially theistic grounds I hold that it is possible for the human mind to work out a physical theory that would fit all aspects of the physi-

[31] I am referring to Paul Davies' *The Mind of God: The Scientific Basis for a Rational World* (New York: Simon and Schuster, 1992).

[32] On Weinberg's book, see my review of if, 'To Awaken from a Dream, Finally!' (1994); reprinted in my *The Limits of a Limitless Science and Other Essays* (Wilmington, DE: ISI Books, 2000), pp. 149-59.

cal universe. After all, to recall a phrase of the Book of Wisdom, the most often quoted phrase of the Bible during the Middle Ages, 'God arranged everything according to measure, and number and weight' (*Wis* 11:20). This means that as long as one views exact science as 'the quantitative study of the quantitative aspects of things in motion', God's arrangement of the material world should appear fully scientific. On finding that final theory man could be rightly proud. Pride unrestrained is not, however, a blessing. For remedy man may take recourse to Gödel's theorem. As Herbert Feigl, a noted philosopher of science but hardly a religious man, once remarked, 'confession (is said to be) good for the soul'.[33] Confession or not, when an error is made the best policy is to admit it promptly. Otherwise, it becomes ever more difficult to do what nowadays is spoken of as 'damage control'. Those who in various ways swallow the universe as if it were a pill, to recall a remark of Robert Louis Stevenson,[34] do much damage, in the long run at least, not only to their own intellectual reputation, but also to the cause of a healthy intellectual climate. Would that reflections on discoveries promote that cause instead of jeopardizing it.

[33] For details, see my Gifford Lectures (note 10 above), p. 414.

[34] Stevenson's phrase is from his 'Crabbed Age and Youth' (1878). Since Stevenson spoke of the solar system, one wonders what simile would he find today when cosmologists play facile games with universes.

TRANSIENT PHENOMENA IN QUANTUM MECHANICS: DIFFRACTION IN TIME

MARCOS MOSHINSKY*

1. *Introduction*

Transient terms [1] are to be expected in a dynamical description of resonance scattering, from the analogy that this description has with the theory of resonant electric circuits [2]. As is well-known, in circuit theory the appearance of resonances in the stationary current is closely related with the transients of the circuit, as the same parameters (resonant frequencies and damping factors) appear in both [3].

The transient terms in a scattering process contain, besides those that could be considered analogous to the electric circuit theory, terms that are related to the time-energy uncertainty relation [4] as quantum mechanics is used in the description of the scattering.

To understand the physical meaning of the transient terms in the resonance scattering process, it seemed of interest to analyze first the transient effects that appear when the propagation of a beam of particles is interrupted. The more complicated phenomenon, where an actual scatterer (represented by a potential or by boundary conditions [5]) is introduced into the beam of incident particles, is briefly discussed in the papers mentioned.

We deal here with the transient terms in the wave function that appear when a shutter is opened. It will be shown in the next section, that when the state of the beam of particles is represented by a wave function satisfying the time-dependent Schroedinger equation, the transient current has a remarkable mathematical similarity with the intensity of light in the Fresnel diffraction by a straight edge [6]. The transient phenomena have therefore been given the name of diffraction in time.

* Member of Instituto de Física UNAM, email: moshi@fisica.unam.mx.

The form of the transient terms of ψ that appear when the shutter is opened, is strongly dependent on the type of wave equation satisfied by ψ. In the present paper, we analyze the transient terms that appear when the ψs satisfy the Schroedinger equation. Only for this equation is there an analogy with the phenomena of optical diffraction, which has to do with the resemblance that the solutions have with those that appear in Sommerfeld's [7] theory of diffraction.

2. *The shutter problem*

The problem we shall discuss in this note is the following: a monochromatic beam of particles of mass $m=1, \hbar=1$, energy $(k^2/2)$, moving parallel to the x-axis is interrupted at $x=0$ by a shutter perpendicular to the beam, as illustrated in Fig. 1. If at $t=0$ the shutter is opened, *what will be the transient particle current observed at a distance x from the shutter?*

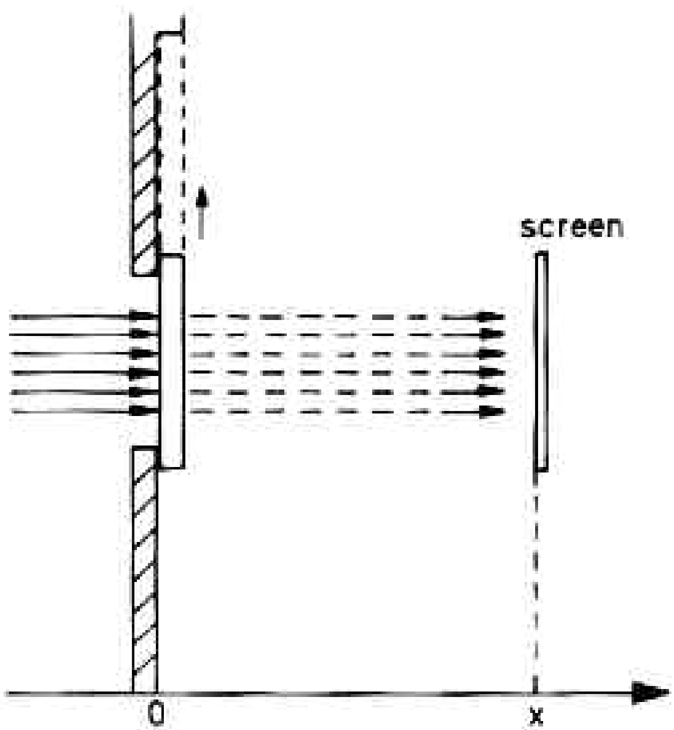


Fig. 1. The shutter problem.

To set the mathematical problem, we must first give the behaviour of the shutter, i.e., if it acts as a perfect absorber (no reflected wave), or a perfect reflector (an infinite potential barrier), or something between the two. For simplicity we will assume that the shutter acts as a perfect absorber, though it can easily be shown that for $x \gg \lambda$ (where λ is the wavelength $\lambda=(2\pi/k)$), the transient current obtained below holds for any type of shutter, so long as it acts as a device that, when closed, keeps the beam of particles on only one side of it.

For non-relativistic particles, the wave function $\psi(x, t)$ that represents the state of the beam of particles for $t>0$, satisfies the time-dependent Schroedinger equation:

$$-i(\delta\psi/\delta t)=(1/2)(\delta^2\psi/\delta x^2), \tag{1}$$

and the initial conditions:

$$\psi(x, 0) = \begin{cases} \exp(ikx), & \text{if } x \leq 0; \\ 0, & \text{if } x > 0. \end{cases} \tag{2}$$

The solution of (1,2) can be given immediately with the help of the onedimensional Schroedinger Green function

$$G(x-x', t)=(2\pi t)^{-1/2}\exp(-i\pi/4)\exp[i(x-x')^2/2t]. \tag{3}$$

We have then

$$\begin{aligned}\psi(x, t) &= \int_{-\infty}^{0} G(x-x', t)\exp(ikx')dx' \\ &= \frac{1}{\sqrt{2}}\exp\left(-i\frac{\pi}{4}\right)\exp\left[i\left(kx-\frac{1}{2}k^2t\right)\right] \\ &\times\int_{-\infty}^{\xi}\exp\left[i\left(\frac{1}{2}\right)u^2\right]du.\end{aligned} \tag{4}$$

where in the last equation we have made the change of variables

$$u = (x'-x)(\pi t)^{-1/2} + k(t/\pi)^{1/2}, \tag{5}$$

$$\xi = (\pi t)^{-1/2}(kt-x). \tag{6}$$

Introducing the Fresnel integrals

$$C(\xi) = \int_0^{\xi} \cos\left(\frac{\pi u^2}{2}\right) du.$$
$$S(\xi) = \int_0^{\xi} \sin\left(\frac{\pi u^2}{2}\right) du. \tag{7}$$

and making use of

$$\int_{-\infty}^{\xi} \exp\left(\frac{i\pi u^2}{2}\right) du = \frac{1}{2}(1+i). \tag{8}$$

we immediately arrive at the result

$$|\psi(x,t)|^2 = \left\{\left[C(\xi)+1/2\right]^2 + \left[S(\xi)+1/2\right]^2\right\}/2. \tag{9}$$

which has the familiar form of the diffraction of a light beam by a semiplane but with the difference that here ξ is the function (6) of position and time.

If we want to have snapshots of the probability density $|\psi(x,t)|^2$ at given instants of time we can make use of the Cornu spiral [8] diagram of Fig. 2.

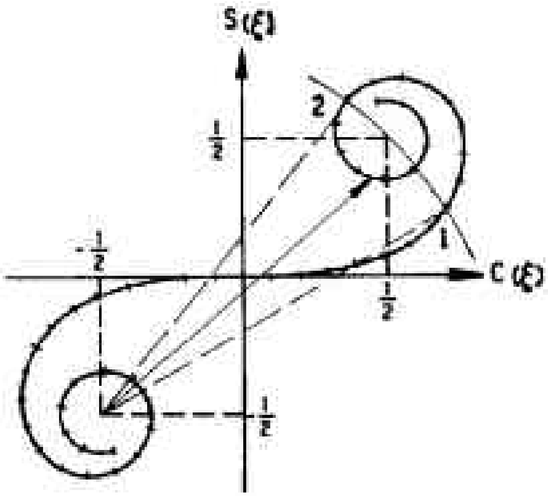


Fig. 2. Cornu spiral [8]. The value of ξ is marked along the curve while the values of the Fresnel integrals $C(\xi)$ and $S(\xi)$ are given along the abscissa and ordinate, respectively. One-half of the square of the magnitude of the vector from the point $(-\frac{1}{2}, -\frac{1}{2})$ to a point on the curve with given ξ gives the probability density for that value of ξ.

The values of $C(\xi)$, $S(\xi)$ are given along abscissa and ordinate while ξ is marked along the curve itself. The value of $|\psi(x,t)|^2$ is one half the square of the distance from the point $(-1/2, -1/2)$ in the plane of Fig. 2 to the given point on the curve corresponding to the value ξ. For t fixed and x going from $-\infty$ to ∞ we see from (6) that ξ goes from $-\infty$ to ∞ passing through $\xi=0$ when $x=x_0=kt$. With the help of the Cornu spiral we then obtain that $|\psi(x,t)|^2$ has the form of the full line in Fig. 3. The classical distribution of particles at time t is indicated by the dashed line terminating abruptly at $x_0=kt$. We indicate in Fig. 3, marking it also with dashes, the probability density at $t=0$ which, from (2), is given by 1 for $x\leq 0$ and 0 for $x>0$.

We see from Fig. 3 that an initial sharp-edged wave packet will move with the classical velocity showing rapid oscillations near the edge. The width of these oscillations can be estimated through the distance $\Delta=x_1-x_2$ marked in the figure between the first two values of x, starting from the edge, in which the probability density takes the classical value 1. The values ξ_1, ξ_2 corresponding to x_1, x_2 are marked in the Cornu spiral of Fig. 2 and from (6) we have

$$\begin{aligned}\xi_1-\xi_2=0.85 &= (\pi t)^{-1/2}\Big[(kt-x_2)-(kt-x_1)\Big]\\ &= (\pi t)^{-1/2}\,\Delta x. \qquad (10)\end{aligned}$$

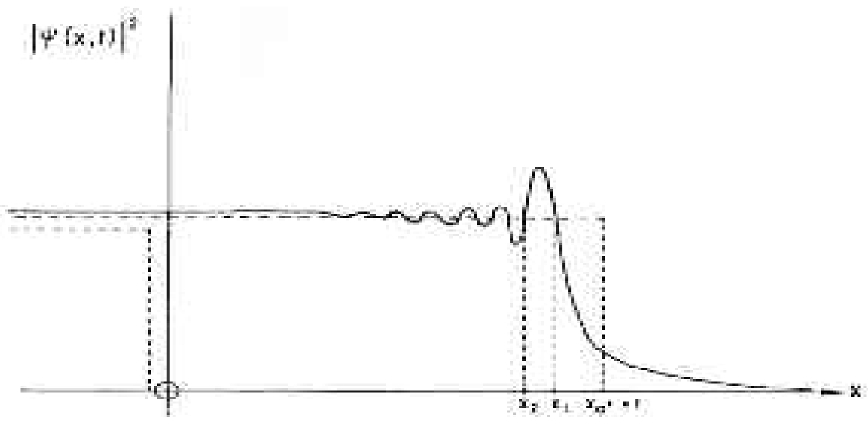


Fig. 3. Probability density of observing a particle at a given time t as a function of distance from the shutter. The dashed line extending to $x=x_0$ represents the classical result at a time t after the opening of the shutter. The dashed line extending to $x=0$ is the probability density before opening the shutter. The values x_1, x_2 are the first positions, starting from x_0, in which the probability density takes its stationary value. In the units we are using $x_0=vt=kt$, where v is the initial velocity of the particle.

Introducing $x_0=kt$ we obtain finally

$$\Delta x = 0.85\ (\pi x_0/k)^{1/2} = 0.85(\lambda x_0/2)^{1/2}, \tag{11}$$

where $\lambda=(2\pi/k)$ is the wavelength of the particle. For particles of wave length $\lambda=10^{-8}$cm at a distance $x_0=10^2$cm from the shutter, the width of the diffraction effect is of the order of 10^{-3}cm.

In Fig. 4 we graph $|\psi(x,t)|^2$ as function of t for fixed x and the width of the resonance in time can be estimated from the difference of the first two times t_2-t_1 at which $|\psi(x,t)|^2$ takes its classical value. This difference can be determined with the help of the Cornu spiral as $\xi_2-\xi_1$ of Eq. (10) continues to be 0.85 and the time width of diffraction in time is, from Eq. (11) given by,

$$\Delta t \equiv \frac{\Delta x}{v} = 0.85\ (\pi x_0/kv^2)^{1/2} \tag{12}$$

where v is the velocity of the particle. If we multiply numerator and denominator by $\hbar$, and use the fact that $\hbar k$ is the momentum of the particle i.e.

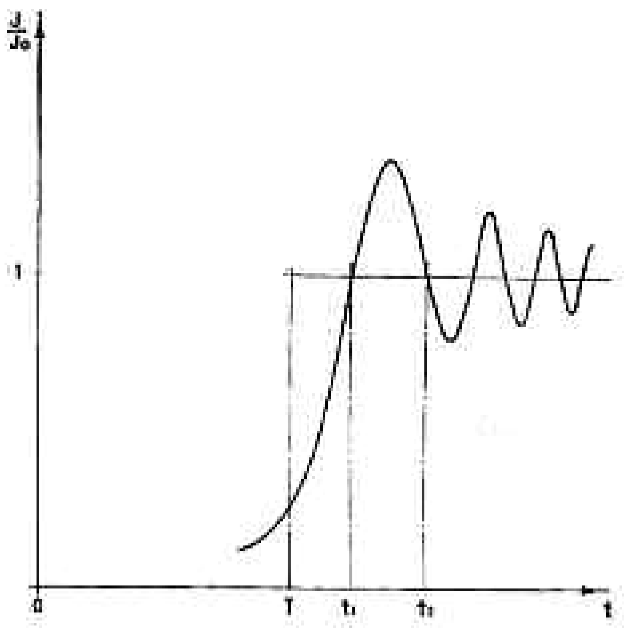


Fig. 4. Width of diffraction in time.

$\hbar k = mv$, with m its mass, we finally have that the width of the diffraction in time is in standard units

$$\Delta t = 0.85\ (\pi x_0 \hbar / m v^3)^{1/2} \tag{13}$$

For $x_0 = 1m$ and neutrons with a velocity corresponding to 300°K the diffraction width is

$$\Delta t = 0.27 \times 10^{-8} \mathrm{sec} \tag{14}$$

If we graph the current as a function of time for a fixed x_0 as shown in Fig. 4 we note that the transient current increases monotically from the very moment in which the shutter is opened and therefore, in principle, an observer at a distance x_0 from the shutter could detect particles before a time x_0/c where c is the velocity of light. This would imply that some of the particles travel with velocities larger than c, and the error is due, of course, to employing the non-relativistic Schroedinger equation in the analysis.

In another publication we employed the ordinary and the Klein-Gordon equation and in both cases there is, at the point x_0, no current observed before the time (x_0/c).

The transient effects associated with the sudden opening of the shutter problem was done originally by the author (*Phys. Rev.* 85, 626, 1952) shortly after he got his Ph.D., and probably was the first analysis of transient effects in quantum mechanics. The field has developed enormously in the last 50 years and he would like to mention the work of M. Kleber 'Exact solutions for time dependent phenomena in quantum mechanics', *Physics Reports* 236 (6) 331 (1994), with 150 references and which presents a very complete review of the progress of the field until 1994.

He would like only to mention the fact that in the original discussion the shutter is opened suddenly. This is not physically possible and Kleber considers a case in which it can be opened slowly using the wave equation

$$\left[i\,\frac{\delta}{\delta t} + \frac{1}{2}\,\frac{\delta^2}{\delta x^2} + \frac{k}{\varepsilon t}\,\delta\,(x) \right] \psi\,(x, t) = 0 \tag{15}$$

The current is in units $(\hbar k/m)$ as function of time t in units $(m/\hbar k^2)$ for both slow $(\varepsilon = 10)$ and fast $(\varepsilon = 0.5)$ lowering of the short-range tunnelling barriers and is given in Fig. 5 at $x = 0$. The exact results (full lines) are compared with the sudden opening of the shutter with dashed dotted lines, which closely agree when $\varepsilon = 0.5$, but are quite different in the case when $\varepsilon = 10$.

3. *Analogues with Sommerfeld's diffraction theory*

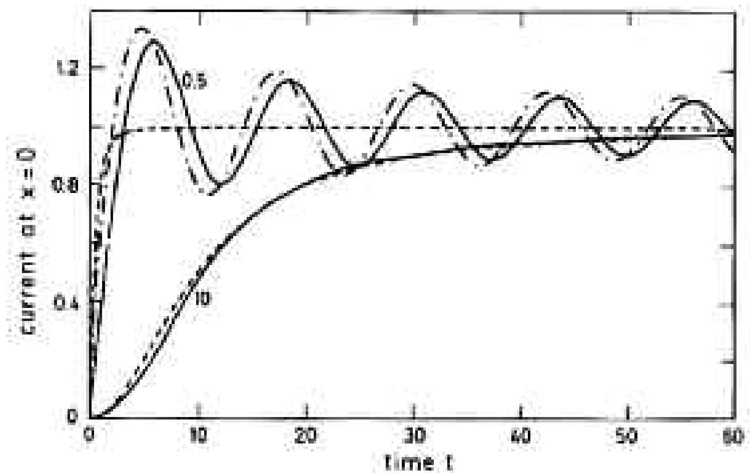


Fig. 5. Tunnelling current (in units of $\hbar k/m$) as a function of time (in units of $\hbar m/(\hbar k)^2$) for slow ($\varepsilon = 10$) and fast ($\varepsilon = 0.5$) lowering of the zero-range tunnelling barrier. The exact results (full lines) are compared with the current derived from the sudden limit (dash-dotted line). M. Kleber, *Physics Report* 236 (6) 1994 p. 384[9].

In the previous sections, the shutter problem for the Schroedinger wave equation was analyzed, and we obtain the corresponding wave functions. Rather than comparing the related transient currents with the intensity of light in optical diffraction [as suggested by the form of the non-relativistic probability density (9)], we shall look into the analogies between the wave functions themselves, and a family of solutions in Sommerfeld's theory of diffraction.

For electromagnetic diffraction problems in the plane, we need appropriate solutions of the two-dimensional Kirchhoff equation, which in polar coordinates has the form

$$(\nabla^2+k^2)\phi \equiv (\delta^2\phi/\delta r^2)+r^{-1}(\delta\phi/\delta r) + r^{-2}(\delta^2\phi/\delta\theta^2)+k^2\phi=0, \quad (16)$$

where $k=(2\pi/\lambda)$ and λ is the wavelength of the radiation.

The well-known family of solutions of (16) that is obtained with the help of the Riemann surface analysis of Sommerfeld [7], can be written in the form:

$$M'(r, \theta, \theta_0) = \exp(ikr)\exp(y'^2)\,\mathrm{erfc}(y'), \tag{17}$$

where:

$$y' = \exp(-i\pi/4)(2kr)^{\frac{1}{2}}\sin\left[\frac{1}{2}(\theta-\theta_0)\right], \tag{18}$$

and θ_0 is a parameter.

The M' satisfy (16) as can be seen from the fact that:

$$\left[\nabla^2+k^2\right]\phi \equiv k^2(2ikr)^{-1}\exp(ikr)$$
$$\times\left[d^2/dy'^2-2y'd/dy'-2\right]\exp(y'^2)\,\mathrm{erfc}(y'), \tag{19}$$

and the right hand side of (19) vanishes.

Despite the fact that the time-dependent Schroedinger equation is of the parabolic type, while the Kirchhoff equation is of the elliptic type, they have solutions of the form (17) though of different arguments.

When a semi-infinite perfectly reflecting plane is introduced at $\theta=-(\pi/2)$, in the path of an electromagnetic wave polarized in the plane and propagating in the direction $\theta=0$, the wave function $\phi(\mathrm{r},\theta)$ that satisfies [7] the boundary conditions at $\theta=-(\pi/2)$, $(3\pi/2)$ and the asymptotic behaviour at $r\to\infty$ becomes:

$$\phi(\mathrm{r},0)=\frac{1}{2}\left[M'(\mathrm{r},\theta,0)-M'(\mathrm{r},\theta,\pi)\right]. \tag{20}$$

From (20) and assuming $x \gg \lambda$ we obtain the characteristic Fresnel diffraction effect [6] for the intensity of the electromagnetic wave in the vicinity of $\theta=0$.

A corresponding problem for the transient current appears when the shutter is represented as a perfect reflector, i.e., an infinite potential barrier.

In this case we note that in reference [10] corresponding to the initial condition [2] the $\psi(x,t)$ of (4) is replaced by

$$\frac{1}{2}M(x,k,t) \equiv \frac{1}{2}\exp(ix^2/2t)\exp(y^2)\,\mathrm{erfc}(y) \tag{21}$$

with

$$y=(-i\pi/4)(2t)^{-1/2}(x-kt). \tag{22}$$

If (2) is replaced $\exp(ikx)-\exp(-ikx)$ we immediately obtain that

$$\psi(x,t)=\frac{1}{2}\left[M(x,k,t)-M(x,-k,t)\right]. \tag{23}$$

From (9) and (23) and assuming also that $kx \gg 1$ we obtain the transient probability density of Fig. 4 which also shows a Fresnel diffraction effect for $t=T$, where T is the time of flight (x/k) in our units $\hbar=m=1$.

REFERENCES

1. Moshinsky, M., *Phys. Rev*., 84, 525 (1951).
2. Wigner, E.P., *Am. J. Phys.*, 17, 99 (1949).
3. Jaeger, J.C., *An Introduction to the Laplace Transformation* (Methuen and Company, London, 1949), p. 31.
4. Moshinsky, M., *Rev. Mex. Fís.*, 1, 28, (1952).
5. Wigner, E.P., *Phys. Rev.*, 70, 15 (1946); Feshbach, Peaslee and Weisskopf, *Phys. Rev*., 71, 145 (1947).
6. Born, M., *Optik* (Julius Springer, Berlin, 1933), pp. 192-5.
7. Sommerfeld, A., *Theorie der Beugung*, chap. XX of the Frank-v. Mises. *Differential and Integralylcichungen der Mechanik und Physik* (Fried. Vieweg and Sohn, Braunschweig, 1935), vol. II, pp. 808-871.
8. Jahnke E. and Emde, F., *Tables of Functions* (Dover Publications, New York, 1945), fourth edition, p. 36.
9. Kleber, M., *Physics Report*, 236 (6) (1994).
10. Moshinsky, M., *Phys. Rev.*, 88, 625 (1952).

OPTICAL METHODS. A SIMPLE WAY TO INTERROGATE AND TO MANIPULATE ATOMS[1]

CLAUDE COHEN-TANNOUDJI

Optical Methods

By letting atoms interact with resonant light, one can
- prepare atoms in interesting states
- detect their presence in these states
- control their various degrees of freedom: spin, position, velocity.

This opens new possibilities, provides new situations, allowing one to explore new fields of research.

Good examples for understanding
- how basic research is making progress
- how important applications emerge, most of the time in an unexpected way, from this progress.

A Few Characteristics of Atomic States

1. Atoms have an internal angular momentum J 

They are like 'spinning tops'.

The projection J_z of J along the z-axis is quantized

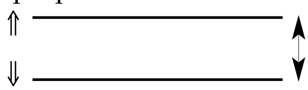
$J_z = M\hbar$ $\qquad \hbar = h/2\pi$ $\qquad$ h: Planck constant

M: magnetic quantum number, integer of half-integer

For example, for a 'spin 1/2' atom, two possible values of M:

$M = +1/2$ Spin up $\Uparrow$ $\qquad M = -1/2$ Spin down $\Downarrow$

In a static magnetic field B, the 2 spin states have opposite magnetic energies proportional to B



Energy splitting ΔE proportional to B

$\Delta E = h\nu_z$

Magnetic resonance: transitions between the 2 spin states induced by a radiofrequency wave with frequency ν_z.

[1] The text below is a transcription of the slides presented during the Symposium.

2. Atoms have a linear momentum P
   $P = mv$ m: mass of the atom v: velocity

A Few Characteristics of Photons

Photons are the light quanta associated with a light beam.

1. Photons have an angular momentum J
   J_z depends on the light polarization.
   Right circular polarization (σ_+) relative to the z-axis: $J_z = +\hbar$
   Left circular polarization (σ_-) relative to the z-axis: $J_z = -\hbar$
   Linear polarization (π) parallel to the z-axis: $J_z = 0$

2. Photons have a linear momentum P
   $P = h\nu/c$ ν: frequency of light c: velocity of light

Basic Conservation Laws

The total angular momentum and the total linear momentum of the atom + photon system are conserved when the atom absorbs or emits a photon.

One can thus transfer angular and linear momentum from photons to atoms, which gives the possibility.

– to polarize the atoms, by preparing them in a given spin state
– to detect (through the polarization of the emitted or absorbed light) in what spin state they are
– to change their velocity.

Using conservation laws in atom-photon interactions for

– manipulating the angular momentum of atoms
  Optical pumping
– reducing their velocity
  Laser cooling and trapping

Optical Pumping

At room temperatures and in low magnetic fields the various atomic spin states are nearly equally populated. Very weak spin polarization.

By transferring angular momentum from polarized photons to atoms, one can achieve large spin polarization and easily detect magnetic resonance signals (due to a tranfer between 2 unequally populated spin states induced by a resonant RF field) in dilute gaseous atomic media

High sensitivity High resolution

Examples of new investigations stimulated by these methods

– multiple photon transitions
– light shifts
– identification of relaxation mechanisms.

A. Kastler

J. Brossel

RADIATIVE FORCES

In a fluorescence cycle (absorption of one photon followed by the spontaneous emission of a photon), the velocity change of the atom due to the transfer of linear momentum of the absorbed photon is equal to $\delta v = h\nu/mc$, on the order of 10^{-2} m/s. The duration of a cycle is limited by the time τ spent by the atom in the excited state before falling back to the ground state by spontaneous emission (radiative lifetime of the order of 10^{-8} s). In a resonant laser beam, the atom can undergo $1/\tau = 10^8$ cycles/s. The velocity change per second, i.e. the acceleration or the deceleration of the atom, can thus be on the order of

$$10^{-2} \times 10^8\,\text{m/s}^2 = 10^6\,\text{m/s}^2 = 10^5\,\text{g}$$

An atom in a resonant laser beam thus feels huge radiative forces which can be used to reduce its velocity (cooling) and to confine it in a small spatial zone (trapping).

Ultracold Atoms

Samples of about 10^8 atoms with temperatures on the order of a few 10^{-6} K and with velocities on the order of 1 cm/s can now be currently obtained. At room temperatures (T = 300 K), atomic velocities are on the order of 1 km/s.

Why are ultracold atoms interesting?

– Low velocities provide long interaction times and consequently allow very precise measurements.

High resolution spectroscopy   Frequency standards

– Low velocities provide large de Broglie wavelengths (proportional to 1/v) and thus make it easier to observe the wave nature of atomic motion

– Interference between de Broglie waves

– Bose Einstein condensation: a macroscopic number of atoms with the same wave function.

Examples Of Applications

– Magnetometers and masers with optically pumped atoms

– MRI of the lung with optically pumped He^3 atoms

– Atomic clocks with ultracold atoms reaching a relative frequency stability and an accuracy of a few 10^{-16}

– Atom lithography

– Atomic gradiometers and gyrometers with de Broglie waves

– Atom lasers: coherent beams of atoms de Broglie waves extracted from a Bose Einstein condensate

– Quantum information using a Bose Einstein condensate trapped in an optical lattice.

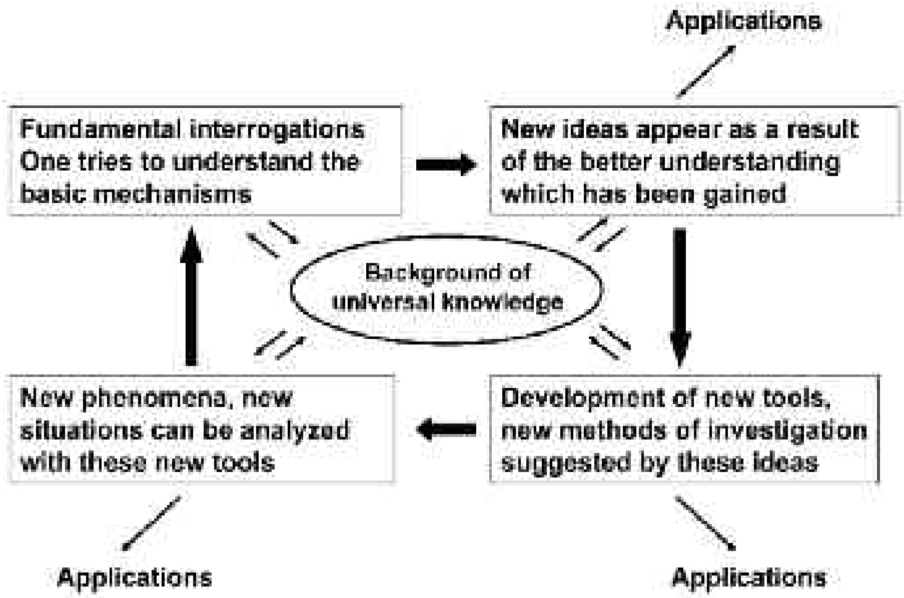
Most of these applications were not planned in advance and introduce discontinuous changes in the technology.

MRI Images Of The Human Chest

Proton-MRI ^{3}He-MRI

G.A. Johnson, L. Hedlund, J. MacFall, Physics World, November 1998 and references therein.

Conclusion



MICRODISCOVERIES: A FRACTAL STORY. A CASE STUDY OF CREATIVE PATHS AND NETWORKS IN SCIENCE

ANTONIO M. BATTRO

SUMMARY

My thesis is that several cognitive processes engaged in any 'microdiscovery' are common to many scientific tasks even up to major discoveries or breakthroughs. I will analyze the paths of discovery of the 'fractal character of saccadic eye movements' as a case study. A power function: $f=k.\ A^{-D}$. relates the number of saccades f with their amplitude A. The exponent is interpreted as the fractal dimension D in the sense of Mandelbrot. Experimental measures show that $1<D<2$. In informational terms $1/D$ could be a measure of the 'temperature of sight'.

I will discuss the importance of 'analogies' in scientific research, the role of 'chance' encounters with relevant information, meetings with people and findings of experimental data. Finally, I will analyze the concepts of innovation, confrontation, modeling, fits and starts, and parsimony in the process of a scientific microdiscovery.

INTRODUCTION

Ainsi lecteur, je suis moy-mesmes la matière de mon livre: ce n'est pas raison que tu employes ton loisir en un subjet si frivole et si vain.
(Michel de Montaigne)

This is a psychological study of a scientific discovery of limited relevance and achieved in a short amount of time, but with a long personal story of several decades of research. It deals with my own mental itinerary in the mathematical and experimental pursuit of the fractal nature of a psychophysical law about eye movements, given as an example. Thus the object of my research is a dual object: (1) a mathematical model (the fractal model) and its experimental application on saccadic eye movements, and (2) a psycho-historical account of the creation of that model based upon my documented recollections simultaneous to the scientific pursuit. Benoît Mandelbrot's famous 'How long is the coast of Britain' (1967) proves that you may increase the length of the coast without limit, provided you take smaller and smaller units of measure. The same could happen here. I am afraid my text could grow without limit if I took smaller and smaller details of my journey toward the discovery of a model for saccadic eye movements. In order to avoid this danger, I will restrict my search only to the first levels of analysis. This essay has the purpose of showing the richness of a standard scientific research, the considerable bulk of associations, images, personal encounters, the complexity of the academic network and the many sheets unfolded during the psychogenesis of a scientific idea. My thesis is that several cognitive processes engaged in any 'microdiscovery' are common to every scientific task right up to the major discoveries. But some important points still remain uncertain. Is the work of genius also of the same kind? Perhaps, if we could describe the common ground of several scientific journeys, the peaks of innovation and creativity will be easily measured against some objective level.

Methodology

Caminante no hay camino, se hace camino al andar.
(Antonio Machado)

I will now describe the methodology I suggest should be followed in a psychogenetic study of discovery.

a) Documentation. I started with the purpose of making short remarks in a log book during the whole process of discovery, and trying to remain objective about places, dates, encounters, ideas, readings, etc. This log has 29 pages in a small (pocket) format, and is handwritten. I carried the notebook with me the whole day during my work. My objective was to become *at the same time* 'the subject and object' of a microdiscovery (a fractal model for saccadic eye movements, in this case) and to enjoy the double *jeu* instead of trying to dismiss one or the other part of it. As a *subject* I was practicing my expertise in eye movements, experimenting, calculating, analyzing data, etc. But, in parallel, I was also the *object* of another experiment, this time in cognition as a *(micro) discoverer* or explorer of some kind, who kept a diary of his journey towards some remote, if any, discovery. In this successful case it took the fractal format of a power function for saccadic eye movements, but many journeys, as we know, even the most simple, may lead to dead ends.

b) Reconstruction. It is essential to have your piece of discovery published in order to 'reconstruct' your own path of discovery, because this is the final text to be confronted as the *terminus ad quem* of your whole endeavor. Once you have finished your scientific task it is time to sit down and retrace your path. You will immediately find that there were several paths and not just one. These paths form a network, but many went loose and some started to diverge. A whole 'landscape of knowledge' will arise under your eyes, even if the final discovery was so tiny as to disappear in the midst of the overwhelming amount of scientific microdiscoveries in your field. But if you retrace your path you will enrich your expertise and your understanding of science in general. And, last but not least, you will offer us a worthy present under the format of a vivid account of an intellectual adventure.

The reconstruction of your path needs some method too. First you might classify the kind or format of the documents as follows:

1. writings, drawings, calculations
2. loose pages or copybooks
3. images and sounds (if videos, records or photographs were taken)
4. different stages of the final text (or different versions, perhaps in different languages)
5. letters to and from the editors, comments of colleagues, referees (including refusals and corrections)
6. conferences, or seminars about your research, before, during and after the publication
7. appearances in media, articles quoting your publication, etc.

Second, classify the 'material contents' of your cognitive journey, and make a catalogue or list of all the items:

1. sources, private and public, remote and recent
2. people, colleagues, friends
3. places, laboratories, libraries
4. trips, fellowships, grants, prizes
5. specific readings and meetings.

Third, write a personal description of the ‘formal contents’, that will include schemata, tables, reflections, etc. about the different cognitive processes (imitation, analogy, deduction, intuition, etc). This is the most difficult part of all the reconstruction, the heart of the matter.

The Sources

Quia minor error in principio magnum est in fine.
(Thomas Aquinas)

My interest in the study of eye movements started as a medical student during my military service in Buenos Aires (1956). I was enrolled as an assistant to the Department of Electro-encephalography of the Hospital Militar Central under the direction of Dr Abraham Mosovich. He taught me how to register ocular movements using common skin electrodes around the eye. The eye works as a dipole (the retina having a negative electric potential of about 1 mV in relation to the cornea) and any ocular movement produces a shift in the electric field of the eye that can be registered by two channels of an EEG. The eye movement angle is a linear function of the potential, for small amplitudes.

After my graduation as a physician (1957) I got a French scholarship (and the next year a fellowship from the University of Buenos Aires) to work with Paul Fraisse at the celebrated Laboratoire de Psychologie Expérimentale et Comparée de la Sorbonne. I worked two full and exciting years with Fraisse in the field of visual perception and I finally presented my thesis for a Doctorat de l’Université de Paris with the title *L’étendue du champ perceptif en fonction du temps d’excitation* (1960). In Paris I used the same technology in electro-oculography that I had learnt in Buenos Aires. I think that the simplicity of the apparatus helped a lot in my research. I published my first paper on eye movements with Fraisse (Battro, Fraisse, 1961) quoted some thirty years later in his book *Des choses et des mots* (1991).

Unexpectedly Jean Piaget made some references to my work in an article related to visual perception in children and adults, published with Vinh Bang in the *Archives de Psychologie* (1961). This was my first academic contact with the great man and it certainly reinforced my enthusiasm for Genetic Epistemology that had already been aroused by the study of *Logique et équilibre* (1957), quoted in my thesis. This book was written by Jean Piaget, Léo Apostel and Benoît Mandelbrot. This was my first contact with Mandelbrot's thinking. A year later I was invited by Piaget to attend his famous weekly seminar at the Centre International d'Epistémologie Génétique in Geneva in 1962. I was then in Switzerland as a student of mathematical logic at the University of Fribourg (I became a fellow of the Centre from 1966 to 1968). In Paris I began to systematically read Piaget who, at that time, published his important volume on perception *Les mécanisms perceptifs* (1961) – too late to be analyzed in my thesis – and back again in Buenos Aires I wrote a *Dictionnaire d'Épistémologie Génetique* (1966, 1972). For my *Dictionnaire* I read the whole work of Piaget and consulted the big corpus of the *Etudes d'Epistémologie Génétique.* The third volume was written by Léo Apostel, Benoît Mandelbrot and Albert Morf with the title: *Logique, langage et théorie de l'information* (1957). In his text 'Linguistique statistique macroscopique' Mandelbrot introduces a 'fraction' 1/B as the index of the 'informational temperature of a text'. It was long before Mandelbrot's impressive breakthrough in Fractal Theory but the idea of measuring the 'temperature' of a collection of discrete events with the help of a fraction became important to me some sixteen years later. I do not remember having studied this fundamental article before, during my stay in Paris, in any case it was not included in the references of my thesis. A long period of latency seems to be a characteristic of many discoveries in science, small or big (Gruber, 1981). And, even more important perhaps is the development in 'fits and starts' of the original idea during an extended period of time.

A Volatile Idea

> *All kinds of reasoning consist in nothing but a comparison, and a discovery of those relations, either constant or inconstant, which two or more objects bear to each other.*
> (David Hume)

In 1972 political and economic reasons forced me to search for work outside my country. Some Brazilian universities accepted my proposal and I began commuting every month by plane from Buenos Aires, where I lived with my wife and three children, to different towns in Brazil. This decade was, in many senses, one of the most productive times of my scientific life. In my monthly air-commuting trips to Brazil during a decade I learned to enjoy working on planes and in airports. I often traveled with a small library of recent books. I was always searching for some new approaches towards visual geometry, a theme I was developing systematically in Brazil with the help of many colleagues. At that time I was engaged in the study and experimental applications of Lie Groups, Riemannian Geometries, Thom Catastrophes, Reggini Visual Perspectives, and the like, in order to understand some new and old perceptual phenomena.

It so happened that on one of those trips we were forced to stay longer at the new Rio de Janeiro airport, on the way back to Buenos Aires. This time I was reading the fascinating book on *Fractals* by Mandelbrot (1975). At the airport I met an old and respected acquaintance, Professor Alberto González Domínguez, one of the most outstanding mathematicians from Argentina. On the plane we did not sit together and at one point I interrupted my friend's dinner to show him the graphics of Perrin's physical Brownian motion and Cauchy's mathematical trails. Mandelbrot wrote about the 'homogenous chaos' of Perrin's graphs. Both highly irregular paths elicited a powerful image in the memory I had of eye movements, in particular the famous A. Yarbus (1967) rendering of the eye scanning the picture of a human face. I compared a Cauchy flight of one of the plates with the saccadic movements I was so used to studying several years before, without hesitation. It turned out that this particular trace was of Fractal Dimension D=1 and it worked as a trigger for the first working analogy between fractals and eye movements. Saccadic movements are also incredibly irregular and their trails amount to a monstrous entanglement during a long visual search. It was natural for me to think of making sense of this essential irregularity. González Domínguez commented: 'Mandelbrot has got a new client!'.

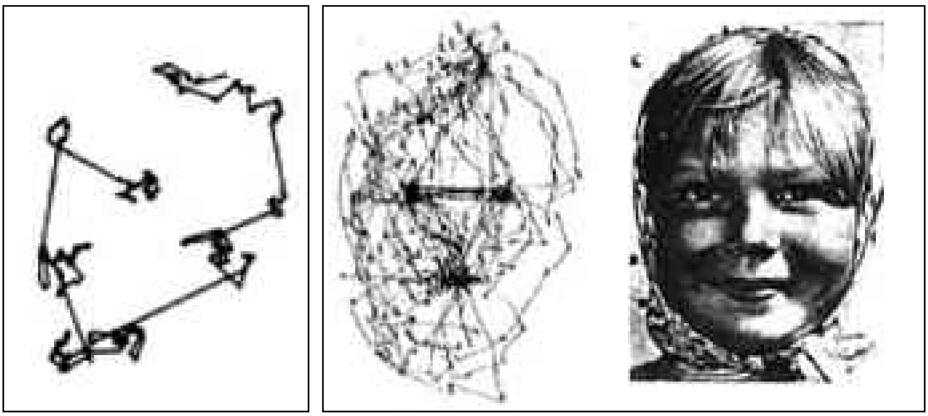


Figure 1. The first analogy: Cauchy flights (left) and eye movements (right) (from Yarbus, 1967).

I must now give some definitions: a *fractal dimension* is a number, not necessarily an integer, that essentially quantifies the degree of irregularity or fragmentation of a geometric set or natural object. A *fractal set* has a fractal dimension that is equal or greater than its ordinary (topological) dimension. A *fractal object* is a natural object that can be represented by a fractal set (Mandelbrot, 1984, p. 154). I rapidly made the hypothesis that the set of saccadic eye movements is a fractal object with a definite fractal dimension D. By analogy with the Cauchy flights on the pictures I postulated a $D \geq 1$ for the saccadic movements. And I decided to prove this statement as soon as possible. In retrospect, the first *visual analogy* between fractals and eye movements that triggered the whole process of my microdiscovery was a plate in a book I was reading several thousand feet above the Atlantic Ocean! It was a mental comparison between some *graphics* on paper and a memory of other graphics. Nothing more. The whole story could have ended there, but this particular visual analogy became the beginning and not the end of a fascinating scientific research.

In 1979 I was very pleased to receive an invitation from Fraisse to spend some months in my former Lab as an Associate Director at the Ecole Pratique des Hautes Etudes. I was surprised and honored by this invitation and I decided to explore the 'fractal connection'. As a preview I gave a seminar at the Centro de Investigaciones Filosóficas, CIF, before my trip, where I had the chance to expose my ideas to a group of colleagues. The CIF is a small private research center founded by a group of friends and myself on my return from Europe in 1965. One of the distinguished visitors to the CIF with whom I shared my fractal ideas was the Argentine mathematician Eduardo Ortiz, then director of the Computer Department at the Imperial College in London. Some months later I paid a visit to Ortiz in London to discuss my fractal problems in some depth.

In Paris I began the search for computerized data with experimental findings on frequency (number) of eye movements as a function of their amplitude (in degrees of an angle). Why? Well, my first reading of Mandelbrot's pre-fractal theories (1956) was in the direction of word distribution, Zipf's law, etc., and I was expecting to find a similar (hyperbolic) distribution in eye movements. When we rank the words in a text by decreasing frequency in a sample of one individual's discourse, the result is a near perfect hyperbolic curve of range/frequency. Moreover, for Mandelbrot (1982, p. 345) 'it is sensible to measure how rich a subject's use of vocabulary is through the relative frequency of his use of rare words'. James Joyce's writings, for instance, have a very high 'temperature of discourse', in informational terms.

I hypothesized that the same process may occur during a visual exploration. A 'warm' gaze, a rich visual scanning of an object, should provide a significant increase of large saccades, a fact that I later interpreted as an increase of the 'temperature of sight'. Hence my first, and desperate, quest for rough data about amplitudes and frequencies. I must say that it was not easy for me to find some reliable eye records. Perhaps my question was not relevant enough for my colleagues. I was stubbornly following my arbitrary analogy between words and eye movements but this was a strange and fantastic guess at this stage. Why this 'hyperbolic obsession'? I was under the impression that if not confirmed by some strong experimental evidence then my whole fractal search could abort, and even my stay in Paris could suffer some strain. It is important to note that the dynamic of saccadic movements is not completely under voluntary control, even if we can choose to fix different points in space, saccades are like 'ballistic' movements. In order to understand the magnitude of the amplitudes in a sample, a movement of 5° corresponds to a scanning of 2.5 cm if we explore a picture at 30 cm from the eye. This small movement takes some 30 ms, larger saccades take more time but at greater speed (i.e. 10°: 5 cm, 40 ms, 15°: 7.5 cm, 50 ms).

Imagine a camera taking a very long record of thousands of ocular saccades in a simple journey in the everyday life of an individual. The result will be a terribly messy path of monstrous geometry, covering an enormous amount of points of the visual field! My search was now engaged in the challenge to 'order' this sort of saccadic chaos. The first result is that if we order the saccades by degrees of amplitude, the result is a nice hyperbolic curve! This fact was already known by some experts but I felt that no one paid much attention to the *theoretical consequences* of this very simple phenomenon. S.S. Stevens, the master of modern psychophysics, firmly established 'the surprising simplicity in sensory metrics' (1962). I understood that the same could be said about saccadic metrics.

In the middle of this search for the fractals I received an invitation from Geneva to discuss a quite different subject, my Brazilian findings about brain laterality and Piagetian operations (Battro, 1981). Incidentally, this was my first contribution to the studies on the brain and cognitive development, that became my central objective two decades later (Battro, 2000). In the train I had the intuition of something to be called 'la température du regard'. The day after Piaget's seminar I returned to my obsessive search. I spent many hours at the University Library reading about Zipf's Law and making acquaintance with Vilfredo

Pareto's powerful insights. I read Zipf (1934-1965) and Pareto's very remarkable *Ecrits sur la courbe de la répartition de la richesse* (1896-1965). I was curious to learn from all these hyperbolic distributions and log-log transforms for words and money! They proved to be of great help on my way to understanding the hyperbolic distribution of eye movements too, at least from the point of view of rough data processing. I felt committed to trying a formula myself following Zipf's and Pareto's examples. That evening, October 30th, 1979, exactly at 7.30 p.m., just before going to dinner with Bärbel Inhelder, Piaget's distinguished colleague, for the first time, I wrote the formula that later became:

$f = k.\ A^{-D}$
f: frequency (number of saccades), A: amplitude of saccades
k: a constant, D: a fractal dimension

I was so excited that I wrote a letter to my family in Buenos Aires, trying to explain my finding and of course I joked about the new 'Battro's Law' at dinner.

Next day I had a meeting with two experimental psychologists, A. Bullinger and J.L. Kauffman at the university. We discussed some technicalities about the 'stability' of eye movements. This time I was not speaking about *length* of saccades (amplitude A) but of *directions* in the visual space (angle). At that point I knew that the number of saccades decreased monotonically as a function of their length (amplitude). But what happened with their directions? I was in trouble because I did not understand the mathematics implied in Mandelbrot's discussion on 'stable distributions' and I did not know how to find some consistent data about the distribution of the direction of saccades in visual space. It was pure chance for me that my colleagues had already obtained remarkable computer star-like graphics showing that the eye can reach any point of the visual field in any direction! When every saccade is geometrically translated to a common origin, a quite regular star is obtained, as in Figure 2.

That suggests that – roughly – there are no privileged angles for saccades. Isotropy implies that every angle (direction of sight) has the same probability. This is of course only true between physiological limits, but in the long run, with thousands of saccades, the star will stretch its rays in all the directions of sight. To me this property seemed essential to the fractal model. As a first approximation the star-like results suggest a stable probabilistic distribution for a large sample of saccades. I am now planning to

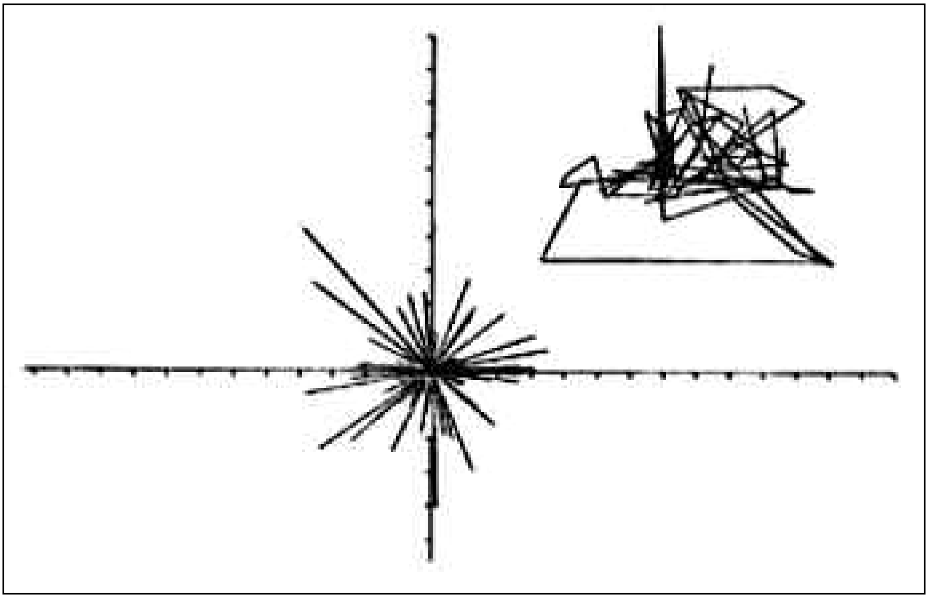


Figure 2. A star of saccades, where eye movements (upper right) are translated to a common origin.

investigate a possible break in this star-like symmetry in some extreme cases of hemianopia (the lack of vision in one half, left or right, of the visual field produced by hemispherectomy; Battro, 2000) more than twenty years after my first observation. A fractal search without end!

That evening, after so much exciting news, I felt very tired. I went to have dinner alone at a fashionable restaurant near my hotel. I dreamt of saccadic stars and power functions. On Thursday night I wrote the first short version of my microdiscovery in French, *La température du regard. Microgenèse d'une idée,* this time at the Café et Restaurant de l'Hôtel de Ville, where I had had so many meetings with friends. Then I went to see 'Le malentendu' the remarkable play by Albert Camus, at the Théâtre de Poche. I wondered if my power function was not also a terrible malentendu. This kind of mixed feelings is very common during the process of discovery, I think. Having reached a harbor, everything has to be put painfully in order again.

Scale and Self-Similarity

Chaque portion de la matière peut être conçue comme un jardin plein de plantes et comme un étang plein de poissons. Mais chaque rameau de la plante, chaque membre de l'animal, chaque goutte de ses humeurs est encore un tel jardin ou un tel étang.
(Leibniz)

The first step of my research was dedicated to understanding the geometric nature of the saccadic paths, below the superficial image of a messy trail of saccades of different amplitudes and orientations, I was lucky to recognize the deep structure of a simple underlying power function. This was an essential step, of the 'empirical' kind, like Zipf's or Pareto's laws for words or salaries. But I also knew that this finding did not suffice. The better I understood Mandelbrot's theory, the more I became aware of the importance of scaling and self-similarity to complete the fractal model of eye movements, but it was difficult, for me at least, to find a ground for this hypothesis. It is obvious that *scaling* plays a quite different role in mathematics, physics and in eye physiology.

Take the Brownian motion of very fine particles (less than 1 micron) as an example. When its motion is examined under the microscope (see Perrin's *Atoms*, 1909, quoted in Mandelbrot 1977), the successive positions can be marked at very small time intervals and joined by segments. The (constructed) prodigious entangled path left behind is a curve of topological dimension DT=1. Mandelbrot says that when a Brownian motion is examined 'increasingly closely' the trajectory increases without bounds and practically fills the whole plane (dimension D=2)! The disparity between these two values DT<D marks the 'fractal nature' of Brownian motion.

The trajectory of saccadic eye movements can also become a monstrous entanglement but it cannot be compared to a Brownian path because of its physiological nature. The search for scaling structures in nature or society is more difficult than in pure mathematics. Below some lower limit the concept of coastline ceases to belong to geography, (Mandelbrot, 1982) and Vilfredo Pareto (1896-1965) also said that his power law 'non vale per gli angioli'. The same for saccadic movements, I understood that the scaling problem should be tackled from another point of view. In order to find some proof I changed from geometry, my first research objective, to Mandelbrot's *lexicographic trees*. This was a subtle shift indeed, but I was guided by the master's hand.

In fact, Mandelbrot, who made the necessary modifications to the Zipf Law in the fifties, also gave some new insights about D as *a similarity dimension* in the field of linguistics in his 1977 book on fractals. Since the Zipf law of word frequency is near perfectly hyperbolic, I quote from his 1982 version, 'it is eminently sensible to try and relate it to some underlying scaling property. As suggested by the notation, the exponent plays the usual role of dimension. An object that could be scaling does indeed exist in the present case: it is a *lexicographical tree*'. This idea proved enough for me, I tried to represent a saccade of amplitude A as a movement between two fixation points #, for instance #aaa# (saccadic amplitude = 3 degrees). I represented this *saccadic tree* as a simple dichotic branching. In the afternoon I returned again to the laboratory and I discussed this idea with the two experts who provided me with the most fascinating computerized eye movements graphs. They told me that some observations suggested a linear continuum from 'micro-saccades' (of minute amplitude) to 'macro-saccades' of the kind I was examining (larger than 1°). I wrote the first English version (4 pages) of my paper at their Lab. The following day I managed to visit the Nestlé Foundation at Lausanne and a remarkable laboratory of architecture at the University where clients could play with 1-1 maquettes of their future homes. Once again I was immersed in the problem of *scale* that became a central issue of my work as a cognitive psychologist of urban space and open places (Battro & Ellis, 1999).

The next days I continued my analysis of the power function, log-log transforms and the like for a while. Zipf based this most general behavior on the 'principle of the least effort'. I wondered at that time whether this 'principle' could explain saccadic eye movements too. Some months later, I received a charming letter from Mandelbrot telling me that he appreciated my findings, but he was not very sure about Zipf's interpretation of the law of least effort. In the last version of my paper I eliminated it.

A Fractal Flight

Das Wandern ist des Müllers Lust, das Wandern!
(Franz Schubert, Wilhelm Müller)

I returned extremely satisfied to Paris after this week in Geneva and decided to find more experimental data on saccade amplitudes. Imagine my joy, and relief, when my colleague Ariane Lévi-Schoen kindly gave me the experimental answer in the format of computerized hyperbolic distributions of saccades as a function of their amplitude! She provided me with 30 nice computer histograms collected in different experiments on vision by her team. The eye movement samples were automatically collected in small intervals of amplitude. I remember that I was disturbed by a little contradictory data for very small amplitudes in some records, but I was relieved when she told me that these cases were sampling artifacts. These beautiful distributions supported my idea that eye movements also follow (as words do) a hyperbolic function (in this case of frequency/amplitude) at least during the limited amounts of time allowed for eye recording in the laboratory. But I was confident that this was a general phenomenon in all natural settings and for all individuals. I was truly excited! I had found what I was looking for after so many months, the experimental data that could fit into a psychophysical model, a general description for a particular complex motor behavior: the hyperbolic function, so common in so many empirical fields! I started to imagine new interpretations following Mandelbrot's ideas about scaling fractals.

With these findings I took the train *La flèche d'argent,* linked to a short trip on an airplane, to London, to meet my friend the mathematician Eduardo Ortiz at the Imperial College. I wanted to better understand 'the probabilistic machine of hyperbolic curves' that was certainly hidden under the saccadic tree. Ortiz was very kind and helpful. We met at South Kensington. He offered to do the computer calculations for me, if needed, in further eye movement experiments. I was excited when I read in J.R. Pierce *Symbols, signals and noise* (1961) that Mandelbrot observed that as a child grows up the power function exponent, in a sample of his vocabulary, decreases from values of around 1.6 to values of around 1.1. I supposed that a similar developmental study could be done for eye movements. Also I planned to analyze the value of D for different settings and individuals, but unfortunately time ran short and I never had the leisure to do it.

I wrote a first version of my work in 1981 with the title: 'La température du regard. Réflexions sur une démarche scientifique', and sent it to Fernando Vidal, then a young Argentine psychologist graduated from Harvard, now a historian of psychology at the Max Plank Institute in Berlin. The first letter from Vidal reported that Mandelbrot had told him that he had already thought about the similarity between saccades and fractals. This was very encouraging for me, indeed! This reminded me of remarkable synchronies in so many macrodiscoveries. But some were quite disturbing, as in the history of non-euclidean geometries. Gauss boldly wrote to Johann Bolyai's father, that he had already discovered what his son tried to prove: 'to praise it would amount to praise myself; for the results to which he has reached, coincide almost exactly with my own meditations'! In London I decided to send Mandelbrot an English version of my fractal model. I wrote him a letter on the train to Oxford.

On my return to Paris I received an invitation to participate in a Symposium on ocular movements at the Maison des Sciences de l'Homme. Several first class international experts were attending the meeting. I understood immediately that I was not on the same 'longueur d'onde'. At one point I asked to make a comment and I stood up in order to draw the hyperbolic curve of the frequency of saccadic amplitudes on the blackboard and explain my idea of a power function and the fractal model underlying it. It was too premature indeed, and I think nobody took any notice of my proposal. Even Mandelbrot at that time was quite unknown to most psychologists in France.

I also received an unexpected invitation from UNESCO to participate at a meeting in Kathmandu. The meeting was about alternative sciences and epistemologies with the grandiose title of *Meeting of experts on philosophical investigation of the conditions for endogenous development of science and technology* (10-14 December 1979). I went to New Delhi via Constance (where I was invited to give a lecture) and Rome. I realized that with so many flights and stopovers, my trip was a fractal path in itself! As a matter of fact the plane was delayed in Rome until the next day, and we were forced to stay at a hotel near Fiumicino. Instead of making a fuss about this delay I remember how happy I was to have more time to write down my notes and memories. I was really enjoying my explicit decision to make detailed annotations about my personal path towards a microdiscovery. I continued to jot down my memories without interruption during the flight to New Delhi, the day after. I still keep the manuscript of 26 pages I wrote in Spanish. With some delay I met my colleagues in Nepal, where I spent some unforgettable days.

On my return to Paris, Fraisse called me at the hotel to invite me to make a presentation about my findings. I still have the schema of my exposé on an envelope. The envelope also had an interesting story. It contained the first computer graphics I was able to obtain of a family of hyperbolic functions and log-log transformations for decreasing exponents D, a very helpful tool indeed for my empirical work on saccadic distributions. Fraisse suggested that I should publish the core of experimental and mathematical findings without the 'cognitive envelope' of the process of discovery. This suggestion revealed the 'duality' of my work, since I analyzed both the object (the fractal nature of saccades) and the subject (my path of discovery) of my research. I submitted my article in the following months under different versions, short and long, to three top scientific journals in turn, without success. Needless to say I was disappointed by the systematic refusals. But now these failures might take on a new meaning in the context of an essay about the psychogenesis of a scientific idea. I have collected the peer reviews concerning my papers. I must say that all of them helped me a lot to explain some critical points of my work better. This is the great advantage of scientific open criticism and I remain very grateful to these anonymous and dedicated experts. Fortunately, Mandelbrot came to my rescue once again and provided me with invaluable moral support. In 1983 he was confident in my fractal model and wrote: 'I regret that your papers were turned down, but this used to happen consistently to my own works before fractals became the rage'.

A Fractal Revival

> *Beware of notions that are new to be sure, but of which no use can be made after they have been defined.*
> (Lebesgue, quoted by Mandelbrot, 1977)

Thanks to the interest of my colleague Miguelina Guirao, the director of the Laboratorio de Investigaciones Sensoriales LIS in Buenos Aires, in 1993 I dug into my old protocols and notes of 1979 to unearth my fractal model of eye movements. She recalculated some experimental samples and obtained a statistical expression of the negative slope of the log-log transform of the data, with a fractal dimension D=1.27 (Figure 3).

As usual, these new and rigorous calculations led me to new insights in the subject matter. I was eager to explore the power of the fractal model and with the help of my young friend Juan P. Garrahan, now a professor of physics at Nottingham University, we produced some simulations of saccadic movements following the two properties, hyperbolic distribution of length of saccades and isotropy of eye movements. Figure 4 is self-explanatory (Battro, 1997).

In June 1994 I returned to Geneva to visit my friends and colleagues, and to attend the Dies Academicus. I returned also to the library in order to search for more data on eye movements. I read the most recent investiga-

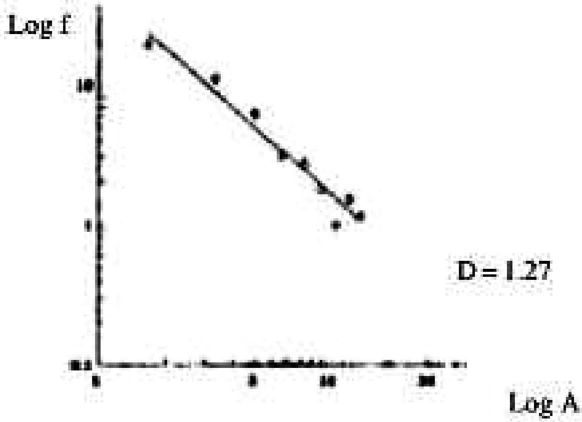


Figure 3. Log-log plot of the experimental data on eye movements. Vertical f: number of saccades, Horizontal: A: amplitude (degrees).

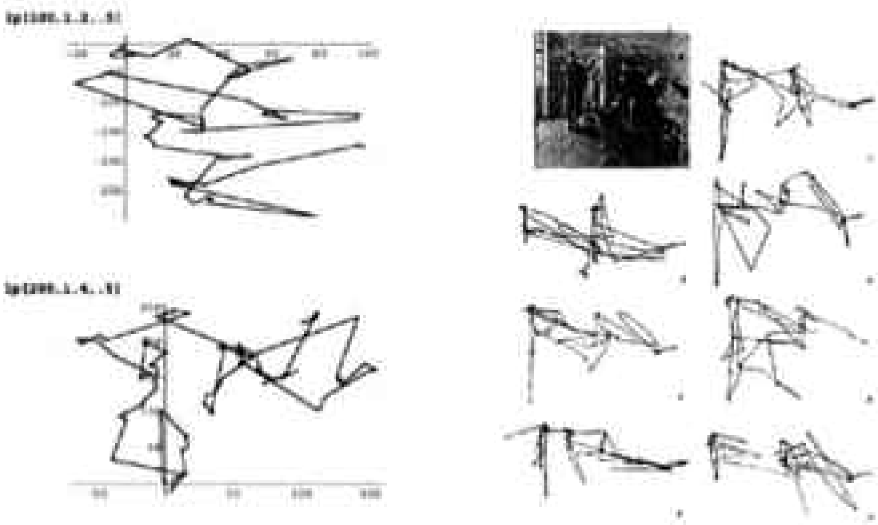


Figure 4. Computer simulation of saccadic eye movements (left). Upper graph: D=1.2, Lower graph: D=1.4. Scanning a picture, from Yarbus (right).

tions without finding any interesting result to add to my old sample on hyperbolic distribution of saccades and very few eye movement records during free search. But I found a most interesting graph in Yarbus' classical book on eye movements. It refers to the microsaccades during fixation on a stationary point. 'The image of the point of fixation always remain inside the fovea… the drift speed *varied chaotically* from zero to approximately 30 minutes of angle per second' (1964). I made a copy of Yarbus' figure 54 and I compared it with Mandelbrot's (1982) figure 255 for a fractional Brown trail of dimensions D~1.11 and D~1.42. As you can see in Figure 5 the traces are remarkably similar in complexity and shape, thus confirming my first vivid impression some decades ago.

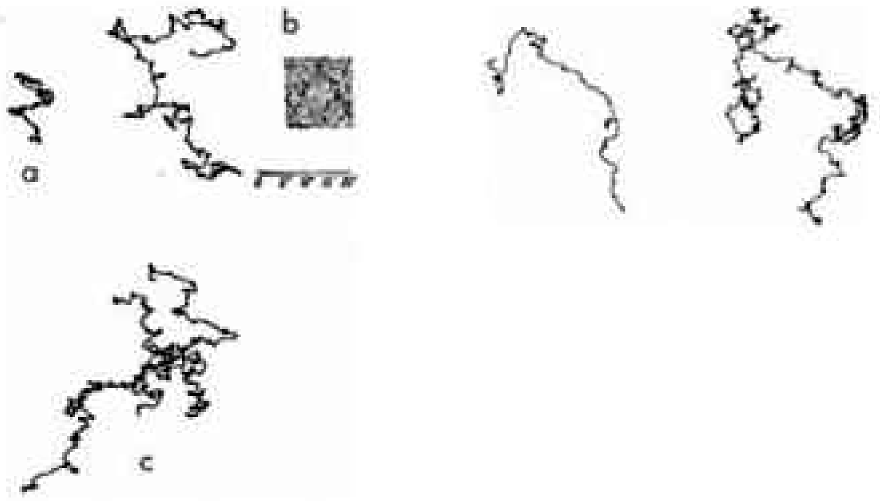


Figure 5. Right: Fractional Brown Trail D~1.11 and D~1.42. Mandelbrot's (1982) Left: three records of eye movements on a stationary point (micro-saccades) a) fixation for 10 sec, b) fixation for 30 sec, c) fixation for 1 min (from Yarbus, 1964).

I discovered two things. First, I returned to my original image, a pure analogy between graphics. This time on a lesser scale, micro-saccades. A new argument perhaps for *fractal space-scaling* (micro-saccades – measured by minutes of angle against macro-saccades – measured by degrees of angle). Second, Yarbus, in retrospect, has given a hint toward the experimental confirmation of *fractal time-scaling* showing three different records for these micro-saccades of 10, 30 and 60 seconds of duration with comparable complexity. Perrin, quoted by Mandelbrot (1982) suggested the property of self-similarity: ‘if the particle positions were marked down 100 times more frequently, each interval would be replaced by a polygon smaller than the whole drawing *but as complicated*, and so on’ (my emphasis). In our case each computer graph of saccades is ‘as complicated’ as the other, for different running times of the simulation.

The Psychogenesis of An Idea

> *From the epistemological point of view we see here how the construction of a new theory is far from being reducible to the accumulation of data, but necessitates an extremely complex structuring of interpretative ideas which are linked to the facts and which enrich them by framing them in context.*
> (Jean Piaget, foreword to H.E. Gruber’s Darwin on Man, 1981)

Now I would like to give some guidelines that might help to follow the genesis of a personal scientific idea:

Agenda: List the people, the places, the letters, the voyages, the drafts, the papers, the conferences, etc. Make a graph with all relevant relations among them. Make a comparison with the story of some important discovery (example: James Watson, *The Double Helix*, 1969). Follow the different paths leading to other microdiscoveries in your scientific life. And try to arrive at some cognitive and epistemological conclusions.

The list of universities where I studied the problem, met relevant people or performed a specific research on fractal eye movements is quite impressive for a modest scientific achievement and it might be compared with some other lists provided by outstanding and mobile scientists during a major discovery. Once again there is a striking commonality between the micro and the macrodiscoveries concerning the number and importance of colleagues and places visited and revisited. I started with a list of 28 people and of the places where I have met them, grouped in 6 classes by the kind of help and expertise (general or specific) they have kindly offered me during my research on eye movements and fractals. It is interesting to add the different personal meeting places related with my research: Buenos Aires, Geneva, Paris, Boston/Cambridge, London. This is good proof of the globalization of science even for a standard research (Suárez-Orozco, Qin-Hillard, 2004). A sociologist of science will also note the predominance of male scientists in my records. I recorded only three women scientists, but their contribution was decisive in many aspects. During this long story five dear masters and colleagues died. Life is broken everywhere at any time, and so are many common and cherished projects. This is a sad fact that must be explicit in every scientific memoir made of flesh and bones. The richer our human network is, the more there are broken threads around us. As for the different documents I conserve on saccadic eye movements and fractals I count 2 manuscripts and 8 papers in different versions (from 1979 to 1994).

And now let us try to analyze 'l'ordre des raisons'. I have found that a substantial part of my research was guided by *analogies*. As a student of philosophy I have read much about the metaphysical importance of analogy. As an epistemologist I was aware of some significant contemporary contributions to the issue, like Mary Hesse's book *Models and analogies in science* (1966). But now, as a scientist, I was confronted by my own insights. My whole research, in fact, was triggered by the analogy between the highly irregular fractal paths discussed by Mandelbrot and saccadic eye movements. Later I developed a kind of mimetic stance with Mandelbrot's formalism, in search of a mathematical expression of the fractal nature of eye movements. I will name the first kind *Image Analogy IA*, and the second *Formal Analogy FA*. I can now summarize the following encounters with analogies on my way to the fractal model:

- IA between self-similarity obtained by the change of scale in fractal natural objects (the measurement of the coast of Britain) and in the construction of scale model maquettes of large or small open places.
- IA between Mandelbrot's mathematical fractal paths (Perrin, Cauchy flights, etc) and eye movements during a visual search (Yarbus).
- FA between some mathematical properties of particular fractal sets, like the hyperbolic distribution in Zipf's and Pareto's laws and the isotropy of space and independence of Cauchy flights.
- FA between the fractal dimension D of a text and the fractal dimension D of a sample of saccadic movements.
- FA between the informational temperature of a literary text and of a sample of saccades (1/D: the temperature of sight).
- FA between Mandelbrot's self-similarity of lexicographic trees and the self-similarity of 'saccadic trees'.
- FA between Perrin's self-similarity in time of Brownian movements and the computer simulation of saccadic movements in short and long run.

I must also underline the central role played by *chance* in my whole research. I propose three main categories: Chance encounters with relevant information (CI), chance meetings with people (CP) and chance findings of experimental data (CE). I can order them in a time-line as:

CI Mandelbrot's pre-fractal analysis of Zipf's law in a volume of the *Etudes d'Epistémologie Génétique* (1957) and my link to Piaget and his Center in Geneva (1962).
CI my first reading about self-similarity and Mandelbrot's fractal objects in *Scientific American* (1978). I used this information in my research about the image of the city and the perception of large open spaces.
CI my first contact with Mandelbrot's fractal flights in the pictures of his book on Fractals.
CP my unexpected meeting in Rio de Janeiro with the mathematician González Domínguez on the same plane where I was reading *Les objects fractales* and the chance to discuss the analogy between fractal paths and saccades with him.
CP Fraisse's invitation to spend some months in Paris as Directeur Associé at his Lab and my desire to explore the fractal world of saccades (1979).
CE the starlike graphs of saccades that were registered for other purposes but to me represented the isotropy of the directions of the eye in the visual space (Geneva, 1979).
CE the hyperbolic distribution of saccades given by experimental histograms (Paris, 1979).

I can end this story of a microdiscovery with an overview of some epistemological relevant properties of my quest.

Innovation: the fractal nature of eye movements is a new interpretation of some experimental facts.

Confrontation: the many versions of my scientific paper, the final paper, lectures and seminars, the personal discussion with the experts, the referees' criticisms, show the amount of interaction caused by my model.

Modeling: the hyperbolic distribution of the amplitude of eye movements allows the mathematical calculation of the fractal dimension D. A computer simulation can show different patterns of saccades for different values of D.

Parsimony: The instruments were restricted to standard equipment and simple calculations and programming. Instead, in many major discoveries new conceptual and/or material instruments are constructed. Perhaps this fact is one of the greatest difference among micro and macrodiscoveries.

Fits and starts: The bulk of the theoretical research took place in only three months, although its roots (and branches) expanded over 47 years (from 1957 to 2004) with some critical bursts of creativity, interspersed with long silent periods. I think this finding is interesting and should lead to comparative studies of the different time-lines in scientific and artistic discoveries. It can be related to some basic and cyclical trends of human cognition, that develops in fits and starts too (Fischer and Rose, 1996). In this particular case I can detect the three following spurts of creativity in my search of the fractal nature of eye movements and I can predict a fourth, starting this year.

I: 1958-60, II: 1979-80, III: 1993. IV: 2004

To sum up, my research on visual perception and eye movements of the period 1958-1960 led my quest to finding some reliable data to fit into Mandelbrot's fractal model in 1979. My frantic search, in only a couple of months, of the hyperbolic distribution of the amplitude of saccades and the isotropic distribution of the directions of eye movements, led to the computer simulations of saccades in 1993. In 2002 Fionn Murtagh, professor of computer sciences at the University of Belfast and Mohsen Farid confirmed other aspects of the fractal nature of the eye gaze, following a complete different discovery path and a very sophisticated technique (Murtagh, Farid, 2002). They also applied eye movements to interface with computers, with what they called the 'eye mouse' (Murtagh *et al*. 2003). Finally, they quoted my work. I was happily surprised and grateful.

REFERENCES

Apostel, L., Mandelbrot, B., Morf, A., *Logique, langage et théorie de l'information* (Paris: Presses Universitaires de France, 1957).

Battro, A.M., *L'étendue du champ perceptif en fonction du temps d'excitation*, thèse d'Université (Université de Paris, 1960), 193 pages, unpublished manuscript.

Battro, A.M. & Fraisse, P., 'Y a-t-il une relation entre la capacité d'appréhension visuelle et les mouvements des yeux?', *Année Psychologique*, 61, 313-323 (1961).

Battro, A.M., *Dictionnaire d'Épistémologie Génetique*, preface by J. Piaget (Paris: Presses Universitaires de France, 1966); *Piaget Dictionary of Terms* (Oxford: Blackwell, 1972).

Battro, A.M., 'Hemispheric lateralization in the development of spatial and logical reasoning in left and right handed children', *Archives de Psychologie*, 49, 83-90 (1981).

Battro, A.M., 'A fractal story: the temperature of sight' (1996), http://www.byd.com.ar/fractalstory.htm.

Battro, A.M., 'La temperatura de la mirada', in M. Guirao (ed.) *Procesos sensoriales y cognitivos. Laboratorio de Investigaciones Sensoriales*, Conicet, (Buenos Aires: Dunken, 1997).

Battro, A.M. and Ellis, E.J., *La imagen de la ciudad en los niños* (1999), http://www.byd.com.ar/ciudad.htm.

Battro, A.M., *Half a Brain is Enough: The story of Nico* (Cambridge: Cambridge University Press, 2000).

Fischer, K.W., Rose, S.P, 'Dynamic growth cycles of brain and cognitive development', in R. Thatcher *et al.* (eds.), *Developmental neuroimaging. Mapping the development of brain and behavior* (New York: Academic Press, 1996).

Fraisse, P., *Des choses et des mots* (Paris: Presses Universitaires de France, 1991).

Gruber, H., *Darwin on Man* (Chicago, IL, University of Chicago Press, 1981).

Gardner, M., 'White and brown music, fractal curves and one-over-f fluctuations', *Scientific American*, April (1978).

Hesse, M., *Models and analogies in science* (Indiana: University of Notre Dame Press, 1966).

Mandelbrot, B.B., 'How long is the coast of Britain? Statistical similarities and fractional dimension', *Science* 155, 636-638 (1967).

Mandelbrot, B.B., *Les objets fractals: forme, hasard et dimension* (Paris: Flammarion, 1975).

Mandelbrot, B.B., *Fractals, form, chance and dimension* (San Francisco: Freeman, 1977).

Mandelbrot, B.B., *The fractal geometry of nature* (San Francisco: Freeman, 1982).

Murtagh, F., Farid, M., 'Statistical properties of eye gaze trace data' (2002), http://main.cs.qub.ac.uk/~fmurtagh/eyemouse/distribution/

Farid, M, Murtagh, M., 'Eye-movements and voice as interface modalities to computer systems', in A. Shearer, F.D. Murtagh, J. Mahon, and P.F. Whelan (eds.), *Opto-Ireland 2002*: *Optical Metrology, Imaging, and Machine Vision*, Proceedings of the SPIE, vol. 4877 (2003), pp. 115-125.

Murtagh, F., Taskaya, T., Contreras, P., Mothe, J. and K. Englmeier, 'Interactive visual user interfaces: a survey', *Artificial Intelligence Review*, 19, 263-283 (2003).

Pareto, V., *Oeuvres complètes* (Genève: Droz, 1896-1965).
Piaget, J., Apostel, L., Mandelbrot. B., *Logique et équilibre* (Paris: Presses Universitaires de France, 1957).
Piaget, J.,Vinh Bang, 'Comparaison des mouvements oculaires et des centrations du regard chez l'enfant et chez l'adulte', *Archives de Psychologie* 38, 150, 167-200 (1961).
Piaget, J., *Les mécanismes perceptifs* (Paris: Presses Universitaires de France, 1960).
Pierce, J.R., *Symbols, signals and noise. The nature and process of communication* (New York: Harper-Collins, 1961).
Stevens, S.S., *Psychophysics* (New York, NY: Wiley, 1975).
Suárez-Orozco, M., Qin-Hillard, D.B., *Globalization: Culture and education in the new millennium* (Berkeley: University of California Press, 2004).
Yarbus, A., *Eye movements and vision* (New York: Plenum, 1967).
Watson, J., *The Double Helix* (New York, NY: Norton, 1969).
Zipf, G.K., *The psychobiology of language* (Cambridge, MA: MIT Press, 1965).
Zipf, G.K., *Human behavior and the principle of least effort* (Cambridge, MA: Addison-Wesley, 1949-1965).

HOW A REFLEXION ON THE INSPIRED NATURE OF THE BIBLE CAN SHED LIGHT ON SCIENTIFIC INVENTION: AN EPISTEMOLOGICAL APPROACH

JEAN-MICHEL MALDAMÉ

This paper proposes to address an important epistemological question. It so happens that its object meets one of the sources of the serious difficulties that mankind is faced with today. As a matter of fact, a number of conflicts are linked to religious attitudes which are grounded on a Text which is considered as Revelation, coming from God or a sacred tradition, and thought to be the only possible source of knowledge. Thus, conflicts of an economical or a territorial nature are overdetermined and carried to excess. And, as I discuss the inspired nature of Biblical texts, I am fully aware that I am approaching a zone of passionate debates.

At an academic level, my reflexion on 'the paths of discovery' encourages me to find, in the religious reference, a motivation for research, and a degree of confidence in the value of scientific work. Its foundation is often found in sacred texts. For some people, an inspired text must coincide with scientific discoveries, while other people consider that the religious field and the scientific field are totally impervious one to the other and are doomed to co-exist without ever merging or combining. Both positions, however, strike me as equally inadequate and unable to account for the unity that exists in the person of the researcher as well as in culture and the history of thought. This is why academics adopt a different viewpoint: that of dialogic complementarity, as 'science and faith are part of a system', which means that they are interacting on each other in a reciprocal process which respects their specific nature.

It is therefore legitimate to try to show how a theological reflexion on new approaches in the religious field can shed light on how scientific invention works – while keeping in mind that, in return, debates internal to science have an impact on theology, since in both cases, what is at stake here is a relation to truth.

1. Intelligence At Work

An image is widespread in the religious world to describe inspiration. It has long been popular in the Christian tradition. It favours the use of a verb, the verb 'to dictate': God is supposed to have dictated a text to a writer; thus, God could be described as the 'author' of the text. All the more so, as the human writer would have been carried away from a normal human condition; the inspired scribe would have been in a trance, described as ecstasy, or rapture (or even 'torpor' – in hebrew, *tardema*).

The use of the verb 'dictate' is based on a certain conception of the relation of God and men, and of nature and grace. For those who are in favour of the conception of inspiration as dictated by God, the message owes nothing to nature, and must be passed on in a manner that breaks away from ordinary daily life. On the other hand, the fact that the inspired person is in a trance allows one to say that his subjectivity has not corrupted the message received from above. Such a conviction is very much present in the traditions which insist on the mediocrity of the messenger, or on his having broken away with the normal conditions of human life. As for Catholic theology, it has given up such a representation. It considers that the writer's activity is important and must be taken into account.

1.1. *The Writer's Activity*

The official position of the Catholic Church rests on two principles, which come under the headings of anthropology and theology.

In the first place, Catholic theology has ackowledged the value of scientific methods applied to Biblical texts. It has introduced scientific methods into the study of texts by using literary, linguistic, historical, archeological agents – and by comparing Biblical texts with non-Biblical texts to study their sources. As a result, it followed that the situation of the Biblical texts was taken into account in the history of universal culture. To acknowledge the structural differences that exist between poetry, epic, historic narrative or a wise man's reflexion, points to an attitude which is not rigid, but well-balanced and humanized. The truth of a text is thus reached through various modes of expression. The taking into consideration of such a diversity of expression and of the writer's historical roots alters the naive notion of inspiration, understood as dictation.

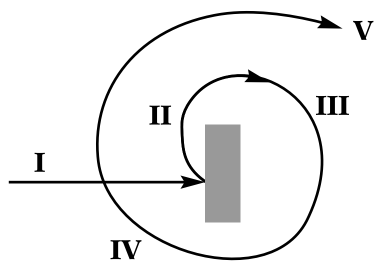
Inspiration should not destroy human subjectivity. To justify everything by resorting to ecstasy amounts to negating the value of the person, of freedom of consciousness, and therefore of human dignity.

The second principle is, properly speaking, of a theological nature, as an answer to the question: what is the underlying conception of God behind the presentation of inspiration as a breaking away from the normal conditions of spiritual life? Would God behave as an enemy of man, or as a wily master? The Catholic tradition firmly rejects such a conception. The document of the Pontifical Biblical Commission on *The Interpretation of the Bible in the Church*, with a preface by Cardinal Ratzinger, acknowledges the value of the scientific approach applied to the Bible. The document recognizes that it is important to tell one literary genre from another, in order to understand the wording of a text.

On these two principles, Catholic theology has made its doctrine of inspiration quite clear. It gives strength to the conception of the action of the Spirit, as formulated by great theologians like Thomas Aquinas and brought up to date by his disciples, among whom Father Lagrange ranks among the greatest.

1.2. *The Prophet in a Tradition*

The most delicate point concerns prophecies. It is indeed easier to acknowledge the labour of a writer who composes a work, of a historian who quotes his sources, or of a jurist who composes a legal code. On the other hand, a prophet is often thought of as the witness of the otherworldly. This is why I shall center my remarks on this point to show that prophecies are the fruit of a labour which demands clear-headedness, efforts and creativeness. When you analyse a prophet's writings, you realize that his resources are those normally at the disposal of a thinker or a man of action, when he tries to express something new. I shall take as a reference a well-known book from the Bible, *the Book of Isaiah*, which plays a major role in the Judaeo-Christian tradition. I shall illustrate it in the following way:



(I) a tradition (II) a blocking point (III) creation (IV) checking (V) opening on to the future.

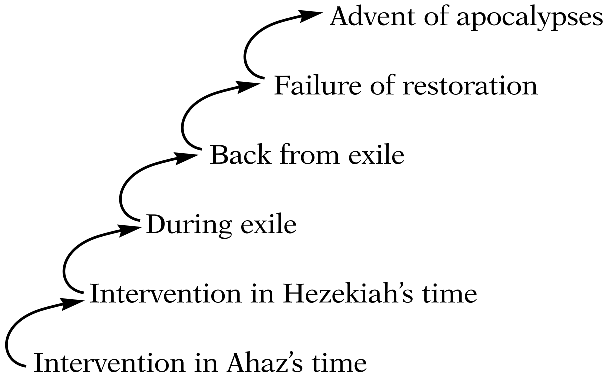
Textual analysis shows that a prophet belongs to a tradition (I). This tradition forms a reference common to a social group: this is why the prophet can rely on it when he addresses his interlocutors. The prophet's mission is to actualise it. Or in other words, he has to show its pertinence and its value at a given moment, when the situation is confused and when those in power do not know what to think (II). Thus Isaiah addresses King Ahaz who despairs of the salvation of Jerusalem, and reminds him of the promise that was made to his father David (*Is* 7, 1-9). In such a somber and dramatic situation (Jerusalem is threatened), the prophet Isaiah explains, clarifies the import of the tradition concerning the city of Jerusalem. His intention is thus inscribed within the framework of a tradition which supports him. By showing the value of such a tradition under new circumstances, he creates something new. The newness lies in his qualities of observation, but also in his skill of expression.The prophet is able to make out, among the noises of the world, what is significant and what makes sense among the phenomena he has observed; he can, thus, produce elements serving to judge the present, and look towards the future.

The words that are pronounced are not only a statement of facts: they point towards the future. Thus, Isaiah says to the King who has just killed his own son, that another son will be born to him and that that son will be a prince of peace (*Is* 7, 10-17). The voice of a prophet is rooted in the past, inscribed in the present, but opens onto the future – thus needing to be verified.

1.3. *The Necessity of Verification*

The Prophet's word is not heard only as the word of God. His conviction does not necessarily entail adherence. So, the word needs to be verified (IV). The Book of Isaiah offers a good criterion of discrimination: is considered as true, the word which is accomplished. When Isaiah says that Jerusalem is the Holy City and that it will not be taken because God resides in it, his word is made true when the siege is raised in a sudden manner. The demand for verification is decisive. His criterion is not an otherworldly behaviour, not some kind of fervour or aesthetic emotion, but the confrontation with facts and events. This is why the prophet himself offers the means of verification – as did Isaiah when, rejected and threatened with death as he was, he decided to confide his word to writing, so that it could be confronted with the coming events (*Is* 8, 16-18).

So, what authenticates the word of the prophet is not the extraordinary, or unusual way in which it is voiced, but the fact that it is firmly rooted in a tradition and offers itself to verification. Isaiah will use the same message when Jerusalem is besieged under King Hezekiah (*Is* 36, 1-37). The above-mentioned diagram repeats itself: this is how a meaningful figure emerges:



This is why it is possible to suggest an epistemological graphic representation for this articulation of historical periods – in order to inscribe what is new in a tradition which opens onto the future. Such a process is universal, and concerns all intellectual approaches of invention. It is of a dynamic nature, and expresses the confidence one can have in the possibility of speaking and making oneself understood. It expresses a specific relationship to time, as mediatized by word.

First stage: receive a tradition

Second stage: find in it a light to enlighten an unsual event

Third stage: verification by fact or events

Fourth stage: adopt an attitude open to the future.

There follows from what precedes a richer approach to knowledge and a widening of one's faith confession. This process of discovery is universal and conveys the principle of Catholic theology, according to which the prophet acts in accordance with the natural capacities of reason and intelligence.

> However divine [the Scriptures] are, they are also human. They bear the traces of human work, of a human composition. On each book, the author has left the mark of his own time and (sometimes very deeply so) of his personality ... It is absolutely not true that the sacred writers had only an easy job to perform. They had to make investigations, to collect materials, to think: not less than ordinary writers (entry: 'Inspiration' in *Dictionnaire de la Bible – Supplément*, Paris: Letouzey & Ané, 1949, col. 482-559).

> They wrote in full possession of their intelligence and faculties, of their own free will, they were really authors, composing their works as authors necessarily do: each according to his own genius and character, each in his own style (*Ibid*., col. 510).

2. An Epistemological Diagram

The preceding comments show how a prophet proposes things that are new. Firmly entrenched in a tradition, on the basis of this tradition, he judges a situation which his contemporaries do not understand. He brings together events, situations and attitudes and shows the links that exist between them. This is how he brings something new into the field of knowledge. There, as is the case in sciences, what counts is a new look at things: this is why prophets are called visionaries, or 'seers' (in hebrew, *nabi*). What is new is of the order of light. The prophet's message is related to reality. One must now therefore examine the relationship that exists between the various fields of knowledge: representation, explanation, interpretation, and understanding.

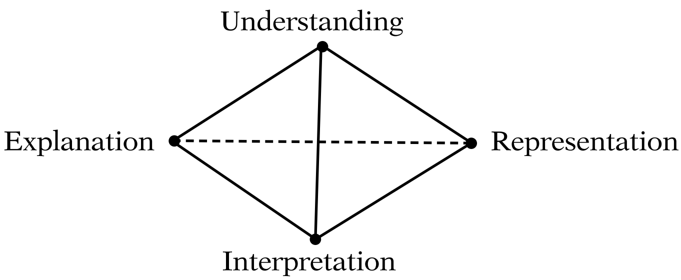
2.1. *A New, Enlightening Answer*

In every activity of knowledge, the discovery of truth goes through four phases: representation, explanation, interpretation, and understanding. It is in the relationship that exists between thoses four phases of knowledge that lies any discovery of something new.

1. In the first place, there is representation. It would be naive to think that a discourse is true, because its terms, notions, or concepts adequately correspond to the structure of reality – as it were by some sort of transfer. In the prophet's language, there are terms, images, relations which voice a certain outlook on things. This is why biblical prophecies have clashed with the way in which kings and priests looked at things. The prophetic word destroys the network of the obvious and establishes connections which no one imagined or thought possible. The word creates a new universe. This is why talking of description is not enough: this is why one must talk of representation.
2. The representation must be linked to an explanation which seeks, behind the phenomenon, an order, a structure, a sequential pattern which refers to something beyond immediate perception, in such a way

that the object of the discourse cannot be identified as the ultimate reality. This is why, in epistemology, one speaks of phenomena.
There is a necessary connexion between representation and explanation, because the prophet knows that events are not the fruits of sheer chance, but actualize a plan – what we call the history of salvation, or Holy history.

3. The passage from representation to explanation is achieved through intelligence. The prophet takes upon himself the responsibility for his assertions; it is the work of interpretation which judges and pronounces itself on the amount of truth which is implied in his analysis and reflexion. This is where the commitment of the prophet lies in the uttered word.
4. But if we consider those three points only (representation, explanation and interpretation), there is a risk of falling into the error of solipsism, or of an infantile form of omnipotence, which genetic psychology describes as magic. There is another demand, for understanding. To understand is to acknowledge that what is taking place is not only a theoretical construction, but has to do with reality, and therefore with a claim to objective truth.



2.2. *Demand for Universality*

The validity of the epistemological pattern is confirmed by the destiny of the prophetic word. The change brought about by the prophet's word is not limited to only one point. It opens a wider perspective. What is new, is not any particular doctrinal point, but a point which is at the core of a whole network of signification.

Thus emerges a totally new way of looking at facts. The prophet proceeds to a fresh reading of history as a whole. He determines, in the course of time, events which he shows are of a founding nature, and cannot be reduced to the normal flow of insignificant events.

Thus, in the second part of the Book of Isaiah (ch.40-55), one can see a theology of the word taking shape and developing, which modifies the way in which his people, till then, perceived the world. Isaiah's theology no longer presents God's action as it were on a small scale. He presents creation as an action through the word, which privileges relationship in full respect of autonomy and gives foundation to the study of the world. The world – created by the word – could not be considered as devoid of sense or of immanent rationality. There is then a justification for knowledge, which recapitulates previous learning and gives it more than a representation or an explanation – something which is a characteristic of understanding. This is confirmed by its fruitfulness – a point which must now be addressed.

2.3. *An Eschatological Horizon*

The way to newness is opened by the discovery of a new way to look at things. It starts in the present, in relation with past and future events. But it is not satified with following the course of time: it is also supported by a demand for understanding which reaches far beyond obstacles. The first obstacle is that of the irrational. Where other people see something absurd and nonsensical, the prophet sees a reason. It refers beyond the realm of explanation, to a more radical sort of intelligibility, which presents itself like a horizon.

The word horizon is characteristic of the fundamental attitude of research. One must map out its various elements.

1. There is a horizon for someone who looks at a distance in a direction which accounts for his walking. Thus, the driving force behind his research is not simply a dissatisfaction with the present, but a desire which motivates the research. To be specific, the prophet's desire is explicitly voiced in terms of a quest for God, in a direction which is that of a tradition which he renews.

Because this approach is that of faith which does not associate with the wording of things but with reality (according to an expression of Thomas Aquinas), it is possible to describe this horizon with one word: that of real.

2. In relation to a certain desire, a horizon is linked with a certain manner of seeing things. Research only progresses by acknowledging one's limitations and specificity.

3. A third point, which is part of the notion of horizon, concerns movement or moving forward. The horizon moves forward as one goes along. This is why the value of a discovery is linked with the way in which it bears fruit. The analysis of the book shows that the prophet Isaiah's word has

been ceaselessly resorted to. His first intervention, during the royal period of King Ahaz, strengthened the movement which later structured itself around a messianic expectation, causing David's city to become the object of a worldwide design. The third part of the Book of Isaiah gives to this messianic expectation a more universal dimension, since in the great city, the capital of a vast empire, the faith of his people was confronted to the riches of a great civilisation.

Thus newness escapes the person who introduced it; it becomes the common good of knowledge. It defines a new field of rationality which remains open to new discoveries.

Conclusion

The pattern displayed in the above chapter shows that the functioning of the prophetic word has nothing singular about it. It has something to do with a human way of relating to truth. A kind of truth that cannot be reduced to a knowledge which has been verified, formulated in a logical manner. This way of connecting the various elements of research together (representation, explanation, interpretation and understanding) is well in keeping with what today's science has recently gone through, in quantum physics as well as in biology, when they gave up resorting to determinism and to the logicism of classical science.

This allows us now to discuss an essential aspect of discovery, where the singularity of the researcher's work appears: the decisive point of invention or of discovery.

3. Invention and Intuition

One point remains to be dealt with: the moment of invention. It is normally named intuition. On this point, many issues are confused. It is important to clarify things, starting from the biblical prophecy – or, in other words, with the word in action. It will be easy to see its relation to other modes of intellectual creation.

3.1. *Symbols*

In the Bible, the prophet is the Man of the Word. The remarks concerning the lives of the prophets show that the word of the prophet, word in action expresses the whole being of the speaking person.

This link, between person and word, falls within the province of what humanities call symbolism. Two elements are meeting here: the act of bringing together is the first meaning of the word symbol. A symbol establishes a link and describes a relation of alliance.

So, the prophet Isaiah gives names to children which are programmes for the future of their people – a future of sorrow (in the name of *Maher-Shalal-Hash-Bas*: swift spoils – imminent looting – *Is* 8, 3) or of happiness (in the name of *Immanou-El*: God with us – *Is* 7, 14)!

The word of the prophet creates a new way of looking at things, where the object which is represented is not separated from all the associations and connections, which make it into a significant object. His word cannot be reduced to words, but refers to a whole system. Now, the symbol links the universal to the individual – and therefore serves intuition, which is the moment of invention.

3.2. *Intuition*

The notion of intuition is one of the most difficult notions that exists. It is used in a variety of ways by philosophical schools, and by scientists – often in too general a sense to be able to serve as a criterion. Or else, it rapidly lapses into psychologism, which avoids considering the very nature of knowledge. Those various approaches have one point in common: the unity of spirit. Intuition can therefore be defined as the realisation of the unity of spirit in action. The study of the prophets shows it: whereas it is difficult to penetrate into the intimacy of a prophet, prophetic acts show him in the words he pronounces. Indeed, the word is made up of multiple variations in the voice; but those cannot be separated from their origin, which testifies to an intention, so that the sounds are gathered together into a meaningful whole. When the intention has taken possession of its object in the act of understanding – as above defined – intuition occurs.

The original unity of the word implies that representations be assumed. Intuition does not give direct access to reality itself, but repeats what has already been given and appears in a new light, in a unified resumption and a new commitment of the spirit. Intuition is never entirely without representations. But it occurs when a link has been found between phenomena that so far had been kept apart. It could not possibly happen in a totally empty mind.

From these remarks concerning prophetic activity, one can easily draw elements to illustrate an epistemology of scientific invention, where intuition is never separated from the demand for objectivity.

3.3. *In Search of Objectivity*

In order to understand the notion of objectivity in the field of prophetic inspiration, one must remember that all prophetic activities are linked to a mission. In order to speak of the fidelity of a prophet, a naive approach resorts to the notion of dictation. As it discarded this notion which is contemptuous for man, Catholic theology has not given up objectivity. It has, quite on the contrary, given objectivity a much less naive foundation. As a matter of fact, a prophet pronounces public words and his message is inscribed in a text.

In a text, there are several elements.The first one is the presentation of a text which corresponds to precise questions, that belong to the realm of description (the actors, and the conditions of their actions). The second one is the very nature of the text in what exegetes call the 'literary genre' – a matter of style and structure. The third one is the repercussion of the text on the reader. This shows that the prophetic text only exists in relation to others; and that the relationship is reciprocal between the prophet and those to whom he addresses his message. Thus, prophetic action attains objectivity in so far as it is not self-oriented, but reaches out to the whole community of believers.

In Catholic theology, it is the community which decides on the inspired character of a text – not the claim of its author. It is the community which receives a text as coming from farther away than its human author through the ordeal of time.

These comments which are situated on the epistemological level, do not deter one from asking questions about the psychological conditions of invention. I must then conclude by examining under what conditions invention is made possible; we shall recognize demands which are linked to the nature of human intelligence.

4. Psychology of Invention

4.1. *Availability*

The first element is availability. This is a delicate point, because it requires a great amount of knowledge and at the same time it needs one to distance oneself from that knowledge, which cannot be considered as sufficient. There is a field which escapes investigation. In order to find a new perspective, one must put into practice what philosophers call 'methodical doubt', and what mystics describe as 'the void'.

The analogies with the mystical experience are enlightening. The mystic person is indeed bent on meeting his God, of Whom he knows that He is above all that exists. He must dismiss all representations which would lead to a kind of rigidity. He must discard from his mind all that seems obvious. The mystics are, for that reason, tormented beings: they are never satisfied with what has already been given to them, because they know that it is not the ultimate gift.

In order to follow such a path, it takes a great degree of availability and a lot of humility. Such was the attitude of the founding masters of science. In their scientific discourse, they have accounted for their discoveries through an analogy with the mystics – an analogy which, as we have seen, was founded on prophetic experience.

4.2. *Maturity*

Although the prophet does belong to a tradition, he is not imprisoned by it. He somewhat distances himself from what he has received.

In fact, prophets have always been at odds with their own culture and environment. Such a situation has no value in itself, but it often is the counterpart of a critical situation. The new appears in a perspective of breaking with tradition. It seems to me that this is only the other side of a situation which is much more important: that of inner unity.

The creation of new things implies an internalization of the practical methods and universal principles of faith. This inner unity is the source of a fecund spiritual life, which enables one to make comparisons – often by analogy, but sometimes just by ordinary metaphors – and to operate differentiations – often through a refinement of concepts and a discrimination between the levels of explanation.

This spiritual quality, it seems to me, can be described as 'maturity'. The word does convey the notion that a life of faith is possible only in a global attitude of one's being. It is a way of being. This is why a prophetic experience can only be appraised through the life quality of him who claims to speak in the name of an absolute which cannot be immediately vouched for or verified.

4.3. *Frailness*

In order to create something new, one needs a certain amount of freedom and consequently one needs to experience a certain frailty, one must break with ready-made certitudes and venture onto unknown grounds. Intelligence, fragile as it is, knows how to recognize, in a vivid, intense way, what is. This is the reason why the autobiographical fragments included in the prophetic books show the torments and the doubts of those who propose their messages. Doubts concerning their own mission is a typical experience for prophets. Doubts often show that the prophet's status is not of his own choice, but has been imposed upon him by a will which he cannot control. Prophets often say that they wish to go back to their native countries, and return to the gentle safety of their homes and serene company of their friends. However, the path lies open before them, and they can only move forward.

These are the universal elements which are valid for all creators and innovators, whether artists or scientists. Prophets are also vividly aware of the shortness of time and of the precariousness of life. They are forced to move forward as though for some urgent purpose.

This awareness is closely related to the fact that they are also vividly aware that the object of their knowledge is inexhaustible, since God lies beyond all representations. The prophet knows that he is facing an infinite which no acquired lore can once and for all circumscribe. If he thought differently, he would not take the trouble to seek answers for new questions. He would content himself with inferring answers from what already exists.

At the end of this third part, it appears that the prophet does deserve his title of 'seer'. He is the one who saw what others did not see at the time when he was speaking. Now the notion of sight is at the same time indicative of dimension and of unity. Vision determines either a wider field (when unknown elements are perceived and explained away) or a deeper one (where what was described can now be explained and even understood); but vision is one, since it is in the same one act that various, irreducible elements are grasped. In this matter, one can see the illuminating influence of a principle which is not only a model of representation, but indeed a light which illuminates whatever comes into its field.

General Conclusion

The above theological discussion implements a certain conception of God. It rests on the conviction that God does not exert violence against human beings. God does not impose upon his prophets messages which would be beyond the scope of their abilities and which they would carry like foreign bodies within themselves. God does not force anyone to do what he would not like to do or even to do something which jeopardizes his self-control.

This is why I must conclude by emphasizing why the Christian tradition speaks of inspiration. The reference to the Spirit of God is here relevant. The word Spirit conveys something which falls within the province of intimacy, of a presence, of discretion. There is in the action of the spirit persuasion, respect and patience, as is conveyed in a celebrated prayer of the liturgy of Whit Sunday, where the Holy Spirit is said to warm up what is cold, to make supple what is rigid, to pacify what is tormented and to straighten up what is distorted.

Lastly, I believe that such a conception of inspiration in the Scriptures results in an exclusion of violence. Because to consider a text as coming directly from God, without important human mediations, is not a procedure of which one religion would have the monopoly. Many are the populations claiming to draw their inspiration from founding texts allegedly dictated by God, and supposed to be the source of the one exclusive truth! Between such religious groups, war is inevitable. As a matter of fact, it has taken place in the past, and is still going on today. One of the criteria of truth of a religious denomination is, it seems to me, the way in which the revelation to which it adheres is lived. It is a useful criterion to distinguish between a sect and a religion. It is one of the major issues concerning peace in the world. Such was my concern when I chose to address this subject in the above discussion.

THE STORY OF NEGATIVE SPECIFIC HEAT

WALTER THIRRING

1. Stable and Unstable Matter

A key property of matter usually encountered in the laboratory is mechanical and thermodynamic stability: an isothermal compression of the system increases the pressure, resulting in a positive isothermal compressibility. Likewise, an extraction of heat at constant volume and particle number leads to a reduction of the temperature T and consequently to a positive specific heat. Formally, stability manifests itself in the extensivity of various thermodynamic quantities such as the internal energy and the entropy. It means that they are proportional to the number of particles N in the system. The Hamiltonian H_N is bounded from below by a constant proportional to N. In statistical mechanics H_N is usually approximated in terms of pairwise-additive two-body potentials $\phi(x_i, x_j)$,

$$H_N = \frac{1}{2m} \sum_{i=1}^{N} \mathbf{p}_i^2 + \sum_{i(<j)} \sum_j \phi(\mathbf{x}_i - \mathbf{x}_j), \tag{1}$$

where xi and pi denote the position and momentum of particle i, and m is the particle mass. If the pair potential ϕ exhibits a significant repulsive core – as is the case for atoms and molecules under ordinary conditions – the system is thermodynamically stable and shows all the properties predicted from kinetic theory. During the approach to equilibrium, macroscopic density and temperature gradients diffuse away, and the system ends up in an equilibrium state, corresponding to a stationary solution of the Boltzmann equation, characterized by a homogeneous distribution in space and a Maxwell-Boltzmann distribution of the momenta.

This situation is totally different for purely negative pair potentials without a repulsive core such as the gravitational potential. The total potential energy becomes proportional to the number of interacting particle pairs and, hence, proportional to N^2. The kinetic energy, however, remains extensive and increases proportionally to N. The familiar thermodynamic limit cannot be performed, in which N and the volume V are increased to infinity keeping the density $\rho = N/V$ constant, and the lower bound proportional to N for the Hamiltonian H_N ceases to exist. Such systems are thermodynamically unstable. Even if one starts out with initial conditions for the particles homogeneous in space and Maxwell-Boltzmann distributed in momentum space, one or possibly more clusters are formed in the course of the natural evolution such that a large percentage of the particles are concentrated in a very narrow region in space. The remaining particles form a diluted atmosphere in which the heavy cluster floats. For a certain range of internal energies, the microcanonical heat capacity of such systems may be negative: an extraction of heat from the system results in an *increase* in the temperature. Microcanonically, the temperature is defined by thermal contact with a small thermometer, whereas the canonical temperature is fixed by contact with a big external heat bath.

The theoretical basis for an understanding of such systems was laid by Thirring and coworkers some time ago [1, 2], and was recently reinvestigated [3]. They studied a cell model which is simple enough to be solved exactly, but which nevertheless exhibits all the surprising and unfamiliar properties of thermodynamic instability. In this model the particles interact with each other via a purely-attractive box-type potential but are totally isolated otherwise, forming a microcanonical system. Then one can show analytically that maximum-entropy equilibrium states exhibit thermodynamic instability. For large internal energies E the system behaves like an ideal gas and has a positive specific heat $C_V = dNk_B/2$, where $d = 2$ or 3 is the dimension of space, k_B is the Boltzmann constant, and N the number of particles. If E is reduced below a threshold close (but not exactly equal) to zero, a single cluster is formed in equilibrium and the specific heat becomes negative. Further decrease of E makes the cluster grow and take in more and more particles, simultaneously diluting the atmosphere. Since the potential energy decreases proportional to $N_c(N_c - 1)/2$, where N_c is the number of particles in the main cluster, the kinetic energy and consequently the temperature increases accordingly. This temperature increase is discussed in more detail in Section 2 and is also demonstrated in a short film produced some time ago [4]. The process of condensing more and more particles into the cluster with decreasing E continues until almost no particles remain in the atmosphere, and the equilibrium temperature $T(E)$ reaches a maximum. If E is reduced even more, no further condensation can take place, and the matter starts behaving normally again: the specific heat becomes positive. The equilibrium predictions of the Thirring model for a two-dimensional system ($d=2$) containing a total of $N=400$ particles are depicted by the smooth curve in Fig. 1, where reduced energy and temperature parameters [5, 6] are defined according to

$$e = 1 + \frac{2E}{N(N-1)\varepsilon} \tag{2}$$

and

$$\theta = \frac{2k_B T(\infty)}{N\varepsilon} \tag{3}$$

respectively. Here and in the following, the time argument ∞ refers to thermodynamic equilibrium, and ε denotes the depth of the negative pair potential. From the definition of the specific heat,

$$C_V = 1/(\partial T/\partial E)_V, \tag{4}$$

it follows that regions with negative slope in Fig. 1 correspond to negative C_V.

In spite of its simplicity and artificiality, this cell model explains qualitatively all the salient features of thermodynamically unstable systems. As examples in astrophysics we mention the gravitational collapse of stars after exhaustion of their nuclear fuel, or the formation of stars and galaxies from interstellar matter. Unfortunately, physically realistic pair potentials are much more difficult to handle theoretically. The gravitational $1/r$-potential, for example, exhibits a singularity at the origin which must be weakened by quantum mechanics and defies a purely classical treatment. Furthermore, it is of long range, which complicates numerical treatment. As a first step towards more realistic systems a simple negative Gaussian pair potential may be used:

$$\phi(\mathbf{x}_i, \mathbf{x}_j) = -\varepsilon e^{-|\mathbf{x}_i - \mathbf{x}_j|^2/\sigma^2}. \tag{5}$$

It is of short range, regular at the origin, and well-suited for computer simulation. It has been used by Compagner, Bruin and Roelse [7] to study the equilibrium properties of a class of unstable systems. They have

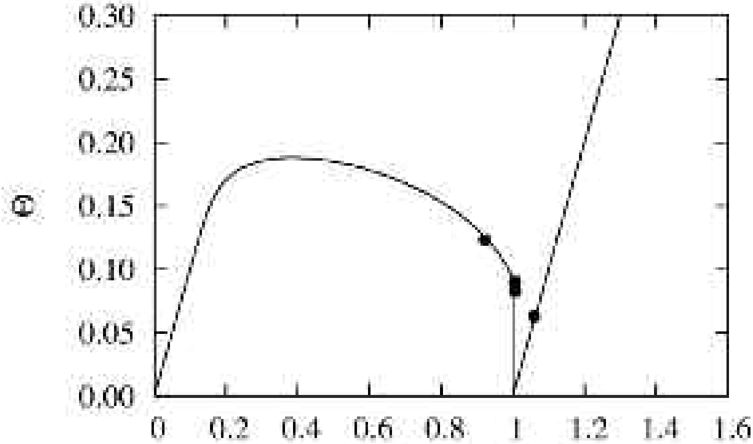


Fig. 1. Plot of the temperature parameter Θ versus the energy parameter e for a 400-particle system in equilibrium. The smooth curve is the prediction of the Thirring model, if the total volume is partitioned into 1600 cells. The points are computer-simulation results for the conditions of the film.

shown that in spite of the differences in potential, the results of the analytical Thirring model above are almost quantitatively confirmed. Using the same pair potential (5), we have performed a series of computer simulations of two-dimensional systems containing $N=400$ particles [5, 6]. Particular emphasis is placed on the transient behaviour as the natural evolution leads the system from a highly non-equilibrium initial state to final equilibrium.

2. Dynamics Of Cluster Growth

In this section some of the dynamical events are described which may be observed in the film mentioned above [4]. Because of the very long relaxation time for equilibration only three characteristic sections of the whole simulation are distinguished below:

1. *Condensation phase*: This regime far from equilibrium covers the early time from the start of the simulation up to a time $t=200$ and shows

the onset of cluster formation. Starting with a homogeneous initial distribution (Fig. 2), a few intermediate clusters are formed very early in the run (Fig. 3). These aggregations compete with each other for further growth. When N_c particles condense into a cluster, the potential energy is locally reduced (i.e. increased in absolute value) by an amount proportional to N_c^2. Since the interaction with the surrounding atmosphere is weak, the local kinetic energy of the cluster has to grow by practically the same amount proportional N_c^2. Thus, the cluster temperature in the major clusters grows proportional to the number of particles N_c in the respective cluster [5, 6]. This behaviour is reflected in the average colour of the particles in the film. With growing N_c the two major clusters turn red indicating a high temperature in the cluster, whereas the surrounding gas still remains at a low temperature. This behaviour very closely resembles the creation of a hot star floating in a cold atmosphere. A quantitative analysis of this property may be found in References 5 and 6.

2. *Intermediate phase*: In this time range $400 < t < 500$ of the simulation the system is still very far from equilibrium. The coexistence of various large clusters in a finite volume as depicted in Fig. 3 does not represent a state of maximum entropy and an equilibrium state [5]. Because of the finite volume V of the system it is inevitable for the major clusters to collide with each other, and a single cluster is formed, thereby significantly increasing the entropy of the system. (Fig. 4). Before the collision the main clusters contain about 40 to 80 particles each. Their combined particle number exceeds the equilibrium value, $N_c(\infty) = 140$, of a single cluster surrounded by an atmosphere [6]. The excess particles are simultaneously evaporated off the final cluster during its birth. This process occurs in the film at a time $t \approx 425$. Because of the dramatic increase of N_c, the temperature in the main cluster jumps to 15, whereas the gas temperature still remains at a low level around 6 as verified from the particle colours.

3. *Equilibrium phase*: This section covering the time range $10000 < t < 10050$ shows the dynamics of the system at equilibrium characterized by a single cluster containing $N_c \approx 140$ particles. The weak interaction between cluster and surrounding gas has been sufficient to raise the gas temperature to the same level as in the cluster. The velocities are Maxwell-Boltzmann distributed both in the cluster and the gas, both distributions being identical [5, 6]. All particles are coloured predominantly red in the film.

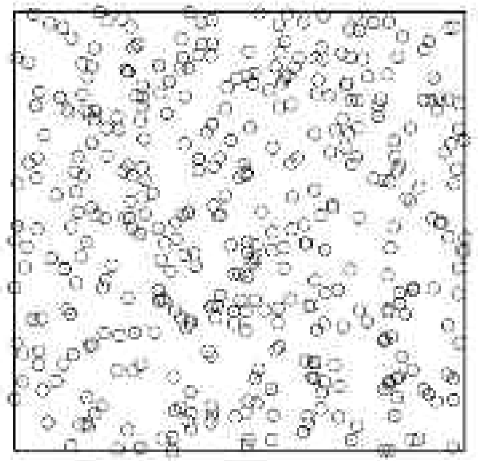


Fig. 2. Random initial configuration. The side length of the central volume is $L = 28.284$.

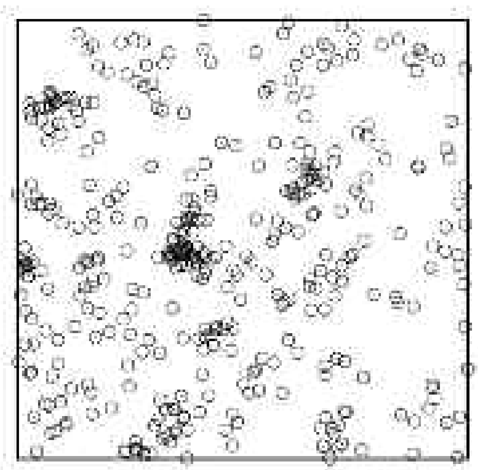


Fig. 3: Instantaneous configuration at a time $t = 20$ corresponding to a state still very far from equilibrium.

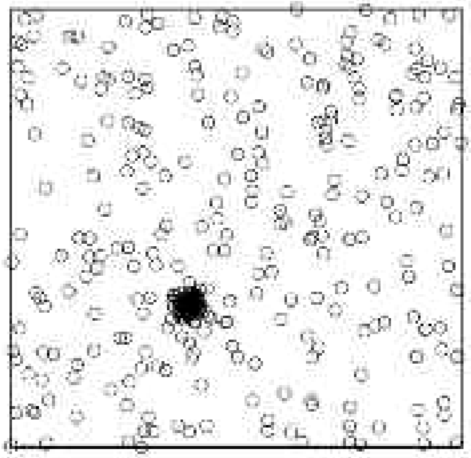


Fig. 4. Instantaneous configuration for the simulation at time $t = 8000$ already close to equilibrium.

It is remarkable that the cluster diameter is practically independent of the number of particles it contains. This fact has been investigated further in Refs. 5 and 6.

In the simulation shown in the film [4], all particles are distributed randomly over the accessible volume V at time $t = 0$. An alternative initial condition would be a state of minimum potential energy $-\varepsilon N(N-1)/2$ for which all N particles are in a perfect cluster and on top of each other. To obtain the same total energy as in the simulation shown in the present film the initial kinetic energy would have to be extremely large. Such a simulation has been carried out independently and reported in Ref. 6. There it is also shown that the initial perfect cluster disintegrates quickly by boiling off particles into the atmosphere. It eventually approaches equilibrium states macroscopically identical to the states shown in the third section of the present film and with the same number of particles $N_c(\infty)$ in a single big cluster.

The most unusual feature of unstable systems is that the microcanonical specific heat may become negative for low enough internal energy E. To understand this phenomenon let us consider for a moment a thermodynamically stable system. Increasing the temperature at constant pressure of the condensed phase, a first order phase transition takes place at a transition temperature T_b. By adding energy, particles are boiled off into the gas at T_b until the condensed phase has disappeared altogether. Similarly, keeping not the pressure but the volume constant any addition of energy leads to an evaporation of particles from the condensed phase to the coexisting gas until the condensed phase has disappeared at a critical temperature T_c. In both cases the field which a condensed particle experiences – and hence its potential energy – is practically independent of the number of particles still remaining in the condensed phase (apart from surface effects). Any increase of the total internal energy E necessarily results in an increase of the kinetic energy of the particles and therefore of the temperature. The specific heat is always positive and the entropy a concave function of E. Since the total energy of stable systems is extensive, the entropy in non-concave regions may always be increased by separating the system into two or more phases, subdividing the total energy accordingly. In such a case the Second Law of thermodynamics tells us to take the concave envelope of the entropies of the individual parts.

For thermodynamically unstable systems, however, any evaporation of cluster particles reduces the field experienced by a particle still in the cluster and increases its potential energy. In such a case the kinetic energy and hence the temperature may become smaller in spite of an increase of the total internal energy E. To understand the phase diagram in Fig. 1, let us first consider the case of an almost totally condensed system for which $e \approx 0$. An increase of e frees only very few particles from the cluster and has only a minor effect on the potential energy of the cluster particles. The temperature also increases and the specific heat is positive as is obvious from Fig. 1. Only after an appreciable amount of cluster particles – say 30% – have been evaporated, any further increase of E results in such a large relative reduction $\Delta N_c/N_c$ of cluster particles that the effect mentioned above is initiated: the temperature is reduced and the specific heat becomes negative. In Fig. 1 this occurs for $e \approx 0.4$. The effect becomes the more pronounced the larger $\Delta N_c/N_c$, and continues to grow until the cluster disappears. This happens for $E \approx 0$ or $e \approx 1$, respectively. However, we cannot speak of a critical temperature in this case because the thermodynamic state of the system is determined by the total internal energy E and not by the temperature. As a consequence, the entropy for thermodynamically unstable systems is not an everywhere concave function of E. This does not contradict the Second Law of thermodynamics because E is not an extensive property for such systems and cannot be subdivided into separate phases. The construction of a concave envelope is of no physical significance.

REFERENCES

1. Thirring, W., *Z.f. Physik*, 235, 339 (1970).
2. Hertel, P. and Thirring, W., *Annals of Physics*, 63, 520 (1971).
3. W. Thirring, H. Narnhofer, and H.A. Posch, *Phys. Rev. Lett.*, 91, 130-601 (2003).
4. Posch, H.A., Majerowitz, A., Thirring, W. and Narnhofer, H., *Condensation phenomena in thermodynamically unstable systems: A simple model for the formation of stars*, Film C 3227, Österreichisches Bundesinstitut für den Wissenschaftlichen Film, Wien, 1990.
5. Posch, H.A., Narnhofer, H. and Thirring, W., 'Computer simulation of collapsing systems', in M. Mareschal (ed.), *Microscopic Simulation of Complex Flows*, Plenum Press, New York, 1990.
6. Posch, H.A., Narnhofer, H. and Thirring, E., *Phys. Rev.*, A 42, 1880 (1990).
7. Compagner, A., Bruin, C. and Roelse, A., *Phys. Rev.*, A 39, 59-89 (1989).

FROM CONTROVERSY TO CONSENSUS IN COSMOLOGY

MARTIN J. REES

1. An Accelerating Universe

Will our universe go on expanding for ever? Will distant galaxies move ever-further away from us? Or could these motions ever reverse, so that the entire firmament eventually recollapses to a 'big crunch'? Until five years ago, the future of our expanding universe was an unsettled question, widely disputed among cosmologists. But progress on several different fronts has allowed us to delineate more confidently what its fate might be.

We used to believe that the answer just depended on the 'competition' between the force of gravity that tends to attract objects to each other, and the expansion energy that tends to move them apart. The Universe will recollapse – gravity will eventually 'defeat' the expansion, unless some other force intervenes – if the density exceeds a definite 'critical' value. This critical density can be computed easily: it corresponds to about 5 atoms in each cubic metre. That does not seem much: but if all the stars in galaxies were dismantled and their material was spread uniformly through space that would amount to only 0.2 atoms per cubic meter – 25 times less than the critical density.

This may seem to imply perpetual expansion, by a wide margin. We have learnt however that the universe contains not just atoms but dark matter – probably some species of particles made in the Big Bang which help to gravitationally bind together stars and galaxies – but even when this is included, the sum seems no more than 0.3 of the critical density. If there were nothing in the universe apart from atoms and dark matter, the expansion would be decelerating, but not enough to ever bring it to a halt.

But observational advances have led to a perplexing twist. A new form of energy – a new force – is latent even in completely empty space. This unexpected force exerts a repulsion that overwhelms the pull of gravity, and causes the cosmic expansion to be speeding up. The evidence came initially from careful study of exploding stars several billion light years away.

Some stars, in their death-throes, explode as supernovae, and for a few days blaze nearly as bright as an entire galaxy containing many billions of ordinary stars. A distinctive type of supernova, technically known as Type 1a, signals a sudden nuclear explosion in the centre of a dying star, when its burnt-out core gets above a particular threshold of mass and goes unstable. It is, in effect, a nuclear bomb with a standard calculable yield. The physics is fairly well understood, but the details need not concern us. What is important is that Type 1a supernovae are pretty close to being 'standard candles' – bright enough to be detected at great distances, and standardised enough in their intrinsic luminosity.

From how bright these supernovae appear, it should be possible to infer reliable distances, and thereby (by measuring the redshift as well) to relate the expansion speed and distance at a past epoch. Cosmologists expected that such measurements would reveal the expected slowdown-rate.

However, the measurements instead provided evidence for an accelerating Universe!

2. Energy Latent in Space: Einstein's Cosmological Constant?

An acceleration in the cosmic expansion implies something remarkable and unexpected about space itself: there must be an extra force that causes a 'cosmic repulsion' even in a vacuum. This force would be indiscernible in the Solar System; nor would it have any significance within our galaxy; but it could overwhelm gravity in the still more rarefied environment of intergalactic space. We normally think of the vacuum as 'nothingness'. But if one removes from a region of interstellar space the few particles that it contains, and even shields it from the radiation passing through it, and cools it to the absolute zero of temperature, the emptiness that is left may still exert some residual force, tending to push things apart.

Einstein himself conjectured this. As early as 1917, soon after he had developed general relativity, he began to think how his theory might apply to the universe. At that time, astronomers only really knew about our own Galaxy – a disc-like swarm of stars; the natural presumption was that the universe was static – neither expanding nor contracting. Einstein found that a universe that was 'set up' static would immediately start to contract, because everything in it attracts everything else. A universe could not persist in a static state unless an extra force counteracted gravity. So he added to his theory a new number which he called the cosmological constant, and denoted by the Greek letter lambda. Einstein's equations then allowed a static universe where, for a suitable value of lambda, a cosmic repulsion exactly balanced gravity. This universe was finite but unbounded: any light beam you transmitted would eventually return and hit the back of your head.

Einstein's reason for inventing lambda has been obsolete for 70 years. But that does not discredit the concept itself. On the contrary, lambda now seems less contrived and 'ad hoc' than Einstein thought it was. Sometimes in science, a new way of looking at a problem 'inverts' it. The question used to be: why should empty space exert a force? Now we ask: why is there not a colossal amount of energy latent in space? Why is the force so small? Empty space, we now realise, is anything but simple. Any particle, together with its antiparticle, can be created by a suitable concentration of energy. On an even tinier scale, empty space may be a seething tangle of strings, manifesting structures in extra dimensions.

All this activity involves energy. Indeed, from this perspective the puzzle is: Why don't all the complicated processes that are going on, even in empty space, have a net effect that is much larger? Why is space not as 'dense' as an atomic nucleus or a neutron star (in which case it would close up on itself within 10-20 kilometres)? Indeed, if lambda represents the energy latent in space, which we realise has intricate structure on subatomic scales, the best theoretical guess is that it should induce a cosmic repulsion 120 powers of ten stronger than is actually claimed!

Most physicists suspected that some process, not yet understood, made the resultant vacuum energy exactly zero, just as other features of our universe – for instance its net electric charge – are believed to be.

But the vacuum energy turns out to be not zero, but it is very, very, small. Why? There clearly has to be some impressive cancellation, but why should this be so precise that it leads to a row of 119 zeros after the decimal point, but not 120 or more? Lambda is remarkably small, but clearly any theory that offers a deep reason why it is exactly zero is wrong too.

A slightly different idea is that the repulsion is due to some all-pervasive field that has negative pressure, and therefore exerts a gravitational repulsion, but which dilutes and decays during the expansion, so that it is by now guaranteed to be very small. This mysterious substance has been dubbed 'quintessence' or dark energy. Yet another possibility, of course, is that Einstein's equations might need modification for some unsuspected reason on the scale of the entire cosmos.

(Just one technical comment. If there is energy in empty space [equivalent, as Einstein taught us, to mass, through his famous equation $E=mc^2$], why does it have the opposite effect on the cosmic expansion from the atoms, the radiation and the dark matter, all of which tend to slow down the expansion? The answer depends on a feature of Einstein's theory that is far from intuitive: gravity, according to the equations of general relativity, depends not just on energy [and mass] but on pressure as well. And a generic feature of the vacuum is that if its energy is positive, then its pressure is negative [in other words, it has a 'tension', like stretched elastic]. The net effect of vacuum energy is then to accelerate the cosmic expansion. It has got a negative pressure and so, according to Einstein's equations, it pushes rather than pulls).

Other independent evidence now supports the case for dark energy. This comes from measurements of the faint microwave radiation that pervades all of space – the afterglow of the hot dense beginning.

Starting out hot and dense, this radiation has cooled and diluted; its wavelength has stretched bringing it into the microwave band. The temperature of this radiation which is lingering on today is only 3 degrees above absolute zero, and it fills the entire Universe. This background radiation is not completely uniform across the sky: there is a slight patchiness in the temperature, caused by the ripples that evolve into galaxies and clusters.

The WMAP spacecraft. is a million miles away, at the Lagrangian point beyond the moon. It is a marvellously sensitive instrument, conceived and designed by a group led by David Wilkinson at Princeton. Wilkinson died in 2002, but fortunately lived long enough to learn of the project's success and to see the early data. It scans the microwave background precisely enough to detect differences of a few micro-degrees between the temperatures in different directions. These variations are imprinted by the precursors of the structures like galaxies and clusters that we see today.

Theory tells us that the most conspicuous waves in the universe – those with the biggest amplitude – are concentrated at particular wavelengths.

There are, as it were, 'resonances' at particular frequencies, just as music in a small room can be distorted by the way particular notes resonate. We can calculate the wavelength of the dominant ripples. But how large they appear on the sky – whether, for instance, they are one degree across or only half a degree across – depends on the geometry of the universe, which in turn depends on the total content of the universe.

If there were nothing in the universe apart from atoms and dark matter with 0.3 of the critical density, the geometry would be what mathematicians call hyperbolic – if you draw a triangle on a saddle-shaped surface, its three angles add up to less than 180 degrees. Light rays travelling across the universe would behave as in a diverging lens; distant objects would then look smaller than they do in Euclidean space where the angles of a triangle add up exactly to 180 degrees.

The WMAP observations have pinned down the angular scale of this amplitude peak: it lies within a few percent of where it ought to be if the Universe were flat. If there were nothing else in the universe beyond the dark matter, we would expect an angle smaller by a factor of 2 – definitely in conflict with observations.

3. Some New Controversies

Cosmology used to be a subject with few facts. But cosmologists now speak of 'concordance'. We now have a convincingly-established framework for interpreting observations of distant objects, and for modelling how the fluctuations in the early universe evolve into the first gravitationally-bound cosmic structures, within which galaxies, stars and planets eventually emerge.

There is indeed remarkable consistency between several independent methods of measuring the key numbers describing our universe. It seems that the universe is flat – in harmony with theoretical prejudices. But there is a new and quite unexpected puzzle. Our universe contains an arbitrary-seeming mix of strange ingredients. Ordinary atoms (baryons), in stars, nebulae, and diffuse intergalactic gas, provide just 4 percent of the mass-energy; dark matter provides 23 percent; and dark energy the rest (i.e. 73 percent). These at least are the values that fit the data best. The expansion accelerates because dark energy (with negative pressure) is the dominant constituent. Of the atoms in the universe, only about half are in galaxies and the rest are diffusely spread through intergalactic space. The most conspicuous things in the cosmos, the stars and glowing gas in galaxies, are less than 2 percent of the universe's total budget of mass-energy – an extraordinary turnaround from what would have been the natural presumption at the start of the 20th century.

There is also firmer evidence for a hot dense 'beginning'. The extrapolation back to a stage when the Universe had been expanding for a few seconds (when the helium formed) deserves to be taken as seriously as, for instance, what geologists or palaeontologists tell us about the early history of our Earth: their inferences are just as indirect (and less quantitative). Several discoveries might have been made over the last thirty years, which would have invalidated the hypothesis, but these have not been made – the big bang theory has lived dangerously for decades, and survived.

But as always in science, each advance brings into sharper focus a new set of perplexing issues: in particular, why was the universe 'set up' expanding the way it is, with such a perplexing mix of ingredients?

Most cosmologists suspect that the uniformity, and the special-seeming expansion rate, are legacies of something remarkable that happened in the first trillionth of a trillionth of a trillionth of a second. The expansion would then have been exponentially accelerated, so that everything in the presently visible part of our universe could have inflated, homogenised, and established the 'fine tuned' balance between gravitational and kinetic energy when that entire domain was only a few centimetres across.

This concept of 'inflation', depending on assumptions about physics far beyond the regime where we have experimental tests, plainly has unsure foundations, but it is not just metaphysics: one can test particular variants of the idea. For instance, the seeds for galaxies and clusters could have been tiny quantum fluctuations, imprinted when the entire universe was of microscopic size, and stretched by inflationary expansion.

The details of the fluctuations depend, in a calculable way, on the physics of 'inflation'. The microwave background, a relic of the pregalactic era, should bear the imprint of these fluctuations. The European Space Agency's Planck/Surveyor spacecraft will, within a few years, yield precise enough data to settle many key questions about cosmology, the early universe, and how galaxies emerged. Such observations will therefore be able to test various possible assumptions about the currently-uncertain physics that prevailed under the extreme conditions of inflation, and thereby at least narrow down the range of options. We will surely learn things about 'grand unified' physics that cannot be directly inferred from ordinary-energy experiments.

4. Beyond 'Our' Big Bang

Another tantalising possibility is that physical reality could be far more extensive than what we have traditionally called our universe. We can only observe the aftermath of 'our' big bang. But there could be an infinity of 'big bangs' within an eternally expanding substratum. Many three-dimensional universes can be packed, so that they do not overlap each other, in a space with 4 or more dimensions. Bugs crawling on a large sheet of paper – their two-dimensional universe – would be unaware of other bugs on a separate sheet of paper. Likewise, we would be unaware of our counterparts on another space-time just a millimetre away, if that millimetre were measured in a 4th spatial dimension, and we are imprisoned in just three.

And the big bangs may all be different. 'Are the laws of physics unique?' is a prosaic paraphrase of Einstein's famous question: 'Did God have any choice when he created the universe?'. Perhaps the ingredients of our universe, and the fluctuations that are the 'seeds' for galaxies, are 'environmental contingencies', imprinted in the immediate aftermath of our big bang, rather than given uniquely by some magic formula. Perhaps, in this enlarged perspective, what we have traditionally called the laws of nature – even those of Einstein and the quantum – could be mere parochial bylaws in our local cosmic patch. There may still be a 'final' theory, at a deeper level, that holds sway over an infinite 'ecological variety' of big bangs.

As an analogy, consider the form of snowflakes. Their ubiquitous six-fold symmetry is a direct consequence of the properties and shape of water molecules. But snowflakes display an immense variety of patterns because each is moulded by its micro-environments: how each flake grows is sensitive to the fortuitous temperature and humidity changes during its growth. The fundamental theory should tell us which aspects of nature are direct consequences of the bedrock theory (just as the symmetrical template of snowflakes is due to the basic structure of a water molecule) and which are (like the distinctive pattern of a particular snowflake) environmental contingencies peculiar to 'our' big bang.

This line of thought is an extrapolation of a perspective-shift that we have already made on a more modest scale – that of stars and planets. We have learnt that millions of stars each have planetary system. So it is unsurprising to find some planets like our rare Earth – planets with the optimum size, temperature, orbit to allow a biosphere. What we have traditionally called 'the universe' may be the outcome of one big bang among many, just as our Solar System is merely one of many planetary systems in the Galaxy. We look for beautiful mathematics in nature, but we do not always find it. Kepler thought that planetary orbits were related by beautiful mathematics. We now know that the Earth's orbit is the outcome of messy and complex dynamics – but happens to end up at a habitable distance from the Sun. The quest for exact formulae for what we normally call the constants of nature may consequently be as doomed as was Kepler's quest for the exact numerology of planetary orbits. And other big bangs will become part of scientific discourse, just as 'other planets' now are.

We still await a 'battle tested' fundamental theory, corroborated by measurements we can actually make, that tells us whether there could have been many 'big bangs' rather than just one, and (if so) how much variety they might display. Until then the epistemological status of the other big bangs is of course precarious.

5. Back to Earth

Humans on Earth are the outcome of four billion years of Darwinian selection. But our solar system is barely middle-aged: the Sun has been shining on the Earth for 4.5 billion years, The unfolding of complexity is just beginning – perhaps humans are still an early stage in this process, rather than (as we sometimes seem to believe) its culmination. There is an unthinking tendency to imagine that humans will be around in 6 billion years, watching the Sun flare up and die. But any life and intelligence that exists then could be as different from us as we are from a bacterium.

Do any general arguments set limits to evolution and complexity – or is the potential literally infinite? The science of complexity probably offers even greater challenges than the 'fundamental' sciences of the cosmos and the microworld.

TOTALLY UNEXPECTED DISCOVERIES: A PERSONAL EXPERIENCE

ANTONINO ZICHICHI

1. The Unexpected Discoveries of the 20th Century

Let me show a synthesis of achievements in Physics during the last Century (Figures 1, 2, 3).

First 50 Years

1 Lorentz Invariance (Lorentz).

2 The quantum nature of the World (Planck).

3 Two worlds: 'Time-like' and 'Space-like' (Einstein).

4 The photon (Einstein).

5 Gauge Invariance (Weyl).

6 The Space-Time curvature (Einstein).

7 Bose-Einstein statistics (Bose-Einstein).

8 The structure of the atom (Bohr).

9 Wave nature of particles (de Broglie).

10 Wave function (Schrödinger) and its probabilistic interpretation (Born).

11 Nuclear Forces (Rutherford) and their 'glue' (Yukawa).

Figure 1.

12 The proof that Quantum Mechanics is self consistent (no contradictions) (Von Neumann).

13 The Weak Forces (Fermi).

14 The Exclusion Principle (Pauli).

15 The Uncertainty Principle (Heisenberg).

16 Fermi statistics (Fermi).

17 The antiparticles (Dirac).

18 The neutron (Chadwick).

19 Time Reversal Invariance (Wigner).

20 Other Invariance Laws (Wigner-Parity; Dirac, Weyl-charge conjugation; Pauli CPT).

21 The neutrino (Pauli, Fermi).

22 The Stars are 'nuclear-fusion' candles (Bethe).

23 Electronic computing (Von Neumann).

24 The sequence of unexpected Fermi discoveries: *Fermi-coupling, Fermi-gas, Fermi-momentum, Fermi-temperature, Fermi-surface, Fermi-statistics, Fermi-transition, Fermi-length* (plus two others quoted above).

Figure 2.

For the first 50 years, I have listed 24 totally unexpected events, with 23 discoveries, and the invention of electronic computers by Von Neumann, which no one could have imagined at the beginning of the 20th Century.

Point no. 24 refers to the impressive list of Fermi discoveries: once again, all totally unexpected.

THE SECOND HALF OF THE 20th CENTURY

25 The Subnuclear World.

26 The Standard Model and Beyond.

27 The Superworld.

Figure 3.

The discoveries of the second 50 years are grouped into 3 classes:

- one is the 'Subnuclear World'
- the second is the 'Standard Model and Beyond'
- the third is the 'Superworld'.

This is the frontier of our knowledge which exists as a fascinating mathematical structure, but lacks the Galilean experimental proof.

The existence of the Subnuclear World and the Standard Model are strictly correlated.

The greatest synthesis of all times in the study of fundamental phenomena (it is called the Standard Model and Beyond [Figures 4, 5, 6, 7, 8]) has been reached through a series of totally unexpected discoveries.

2. The Standard Model and Beyond

The superb synthesis called the 'Standard Model' is a part of a more general structure, where many problems are open. We call this structure 'The Standard Model and Beyond', 'SM&B'.

This Structure brings to the unification of all Fundamental Forces of Nature, suggests the existence of the Superworld and produces the need for a non-point-like description of Physics processes (the so-called Relativistic Quantum String Theory: RQST), thus opening the way to quantize gravity. This is summarised in Figure 5.

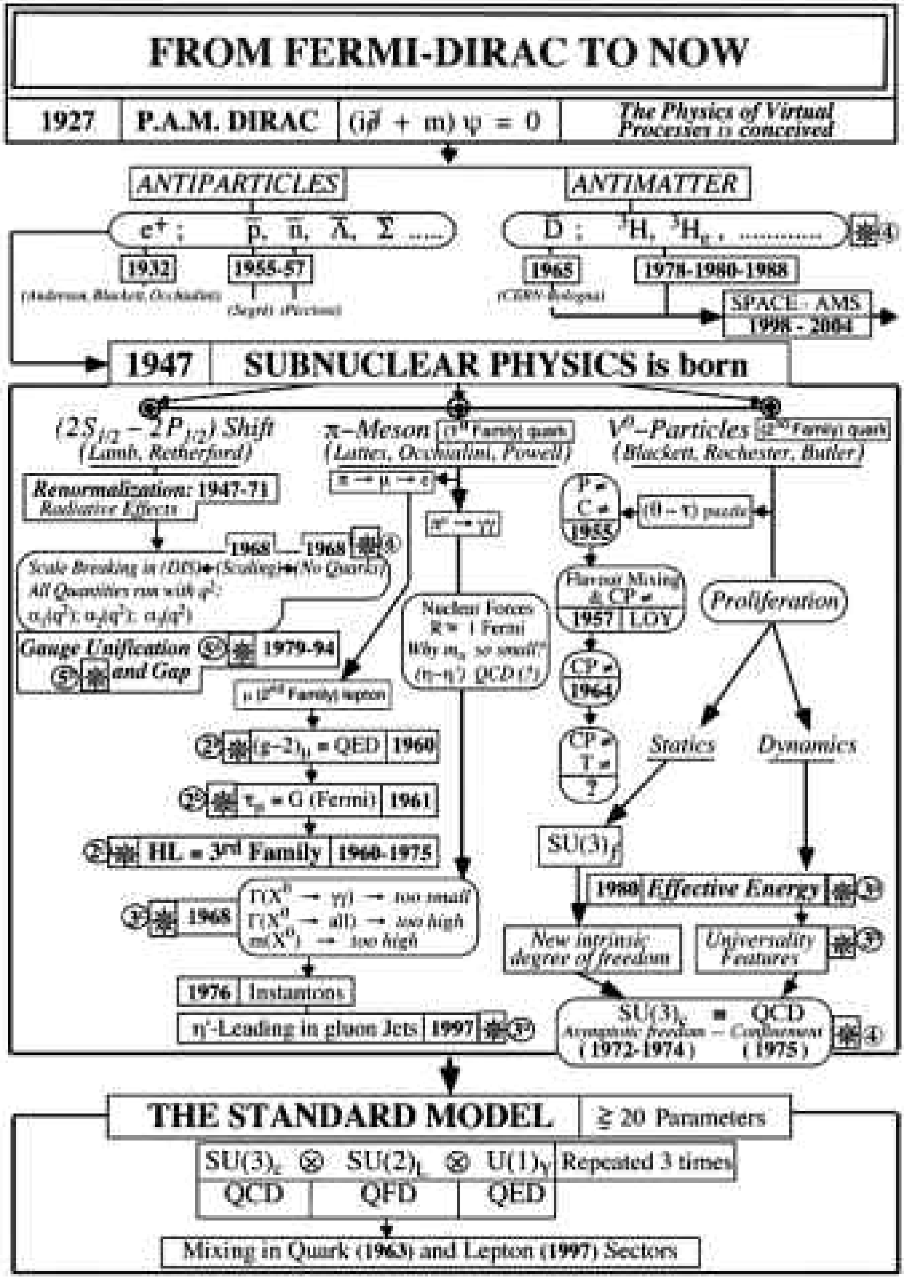

FROM FERMI-DIRAC TO NOW
1927
P.A.M. DIRAC
The Physics of Virtual Processes is conceived
ANTIPARTICLES
ANTIMATTER
1932
1955-57
1965
1978-1980-1988
SPACE - AMS
1998 - 2004
1947
SUBNUCLEAR PHYSICS is born
Renormalization: 1947-71
Radiative Effects
1968
1968
Proliferation
Statics
Dynamics
Gauge Unification and Gap
1979-94
1960
1961
1960-1975
1968
1976 Instantons
1997
1980 Effective Energy
New intrinsic degree of freedom
Universality Features
THE STANDARD MODEL
≧ 20 Parameters
Repeated 3 times
QCD
QFD
QED
Mixing in Quark (1963) and Lepton (1997) Sectors

Figure 4.

DETAILED INFORMATION ON THE STANDARD MODEL AND BEYOND

① RGEs (α_i ($i = 1, 2, 3$); m_j ($j = q, l, G, H$)) : $f(k^2)$.
- GUT ($\alpha_{GUT} \simeq 1/24$) & GAP ($10^{16} - 10^{18}$) GeV.
- SUSY (to stabilize $m_F/m_P \simeq 10^{-17}$).
- RQST (to quantize Gravity).

② Gauge Principle (hidden and expanded dimensions).
— How a Fundamental Force is generated: SU(3); SU(2); U(1) and Gravity.

③ The Physics of Imaginary Masses: SSB.
— The Imaginary Mass in SU(2) × U(1) produces masses ($m_{W^{\pm}}$; m_{Z^0}; m_q; m_l), including $m_\gamma = 0$.
— The Imaginary Mass in SU(5)⇒SU(3)×SU(2)×U(1) or in any higher Symmetry Group (not containing U(1)) ⇒ SU(3) × SU(2) × U(1) produces Monopoles.
— The Imaginary Mass in $SU(3)_c$ generates Confinement.

④ Flavour Mixings & CP ≠ , T ≠ .
— No need for it but it is there.

⑤ Anomalies & Instantons.
— Basic Features of all Non-Abelian Forces.

Note:

q	≡	quark and squark;	m_F	≡	Fermi mass scale;
l	≡	lepton and slepton;	m_P	≡	Planck mass scale;
G	≡	Gauge boson and Gaugino;	k	≡	quadrimomentum;
H	≡	Higgs and Shiggs;	C	≡	Charge Conjugation;
RGEs	≡	Renormalization Group Equations;	P	≡	Parity;
GUT	≡	Grand Unified Theory;	T	≡	Time Reversal;
SUSY	≡	Supersymmetry;	≠	≡	Breakdown of Symmetry Operators.
RQST	≡	Relativistic Quantum String Theory;			
SSB	≡	Spontaneous Symmetry Breaking.			

The five basic steps in our understanding of nature. ① The renormalization group equations (RGEs) imply that the gauge couplings (α_i) and the masses (m_j) all run with k^2. It is this running which allows GUT, suggests SUSY and produces the need for a non point-like description (RQST) of physics processes, thus opening the way to quantize gravity. ② All forces originate in the same way: the gauge principle. ③ Imaginary masses play a central role in describing nature. ④ The mass-eigenstates are mixed when the Fermi forces come in. The Abelian force QED has lost its role of being the guide for all fundamental forces. The non-Abelian gauge forces dominate and have features which are not present in QED.

Figure 5.

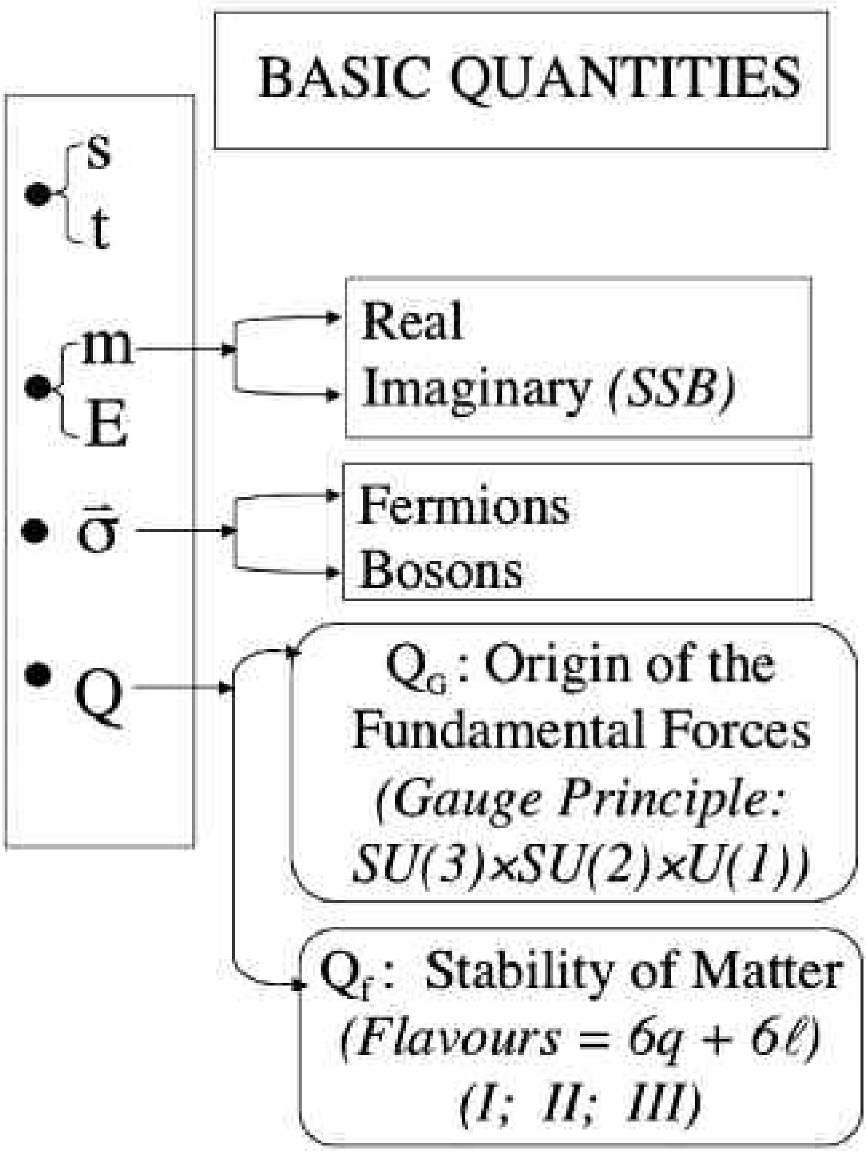

BASIC QUANTITIES
s
t
m
E
σ̄
Q
Real
Imaginary (SSB)
Fermions
Bosons
Q_G : Origin of the Fundamental Forces (Gauge Principle: SU(3)×SU(2)×U(1))
Q_f : Stability of Matter (Flavours = 6q + 6ℓ) (I; II; III)

Figure 6.

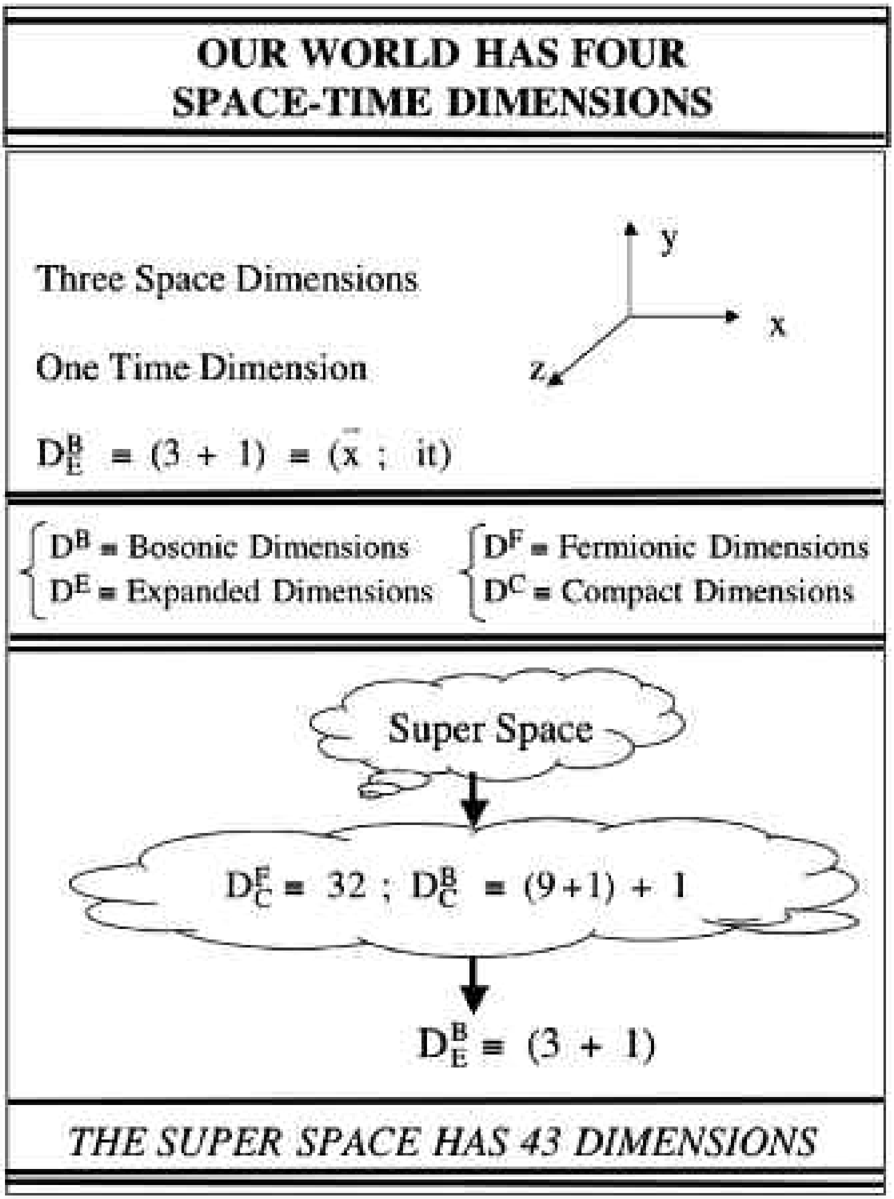

OUR WORLD HAS FOUR
SPACE-TIME DIMENSIONS
Three Space Dimensions
One Time Dimension
y
x
z
$D^B_E \equiv (3+1) \equiv (\vec{x}\ ;\ it)$
$D^B \equiv$ Bosonic Dimensions
$D^E \equiv$ Expanded Dimensions
$D^F \equiv$ Fermionic Dimensions
$D^C \equiv$ Compact Dimensions
Super Space
$D^F_C \equiv 32\ ;\ D^B_C \equiv (9+1)+1$
$D^B_E \equiv (3+1)$
THE SUPER SPACE HAS 43 DIMENSIONS

Figure 7.

3. A Few Examples Where I Have Been Involved

I will now discuss a few cases, where I have been involved (Figure 8).

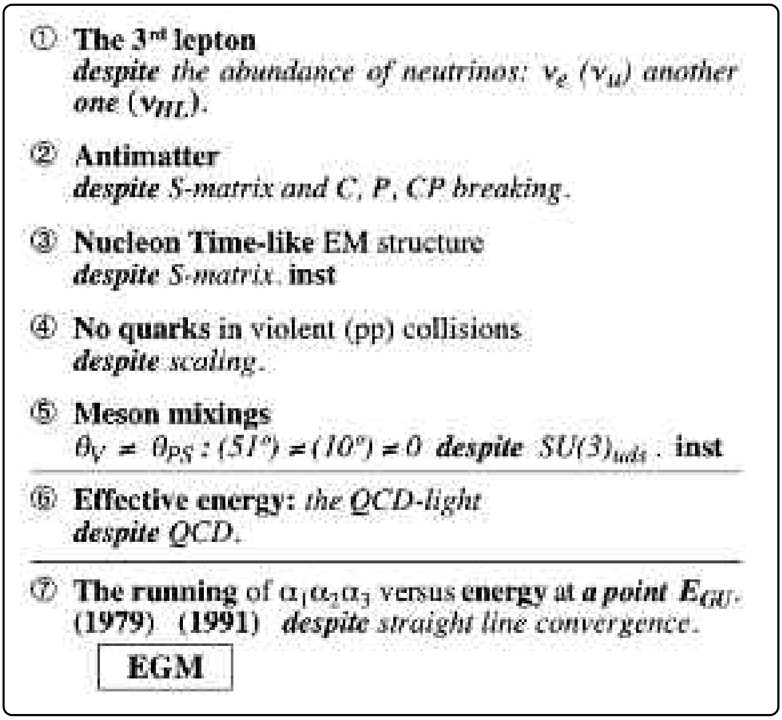


Figure 8.

POINT 1. The Third Lepton, and the other unexpected events in Electroweak Interactions are illustrated in Figure 9.

Note that for the Electroweak force, Nature has not chosen the simplest way out SU(2), but unexpectedly SU(2)×U(1).

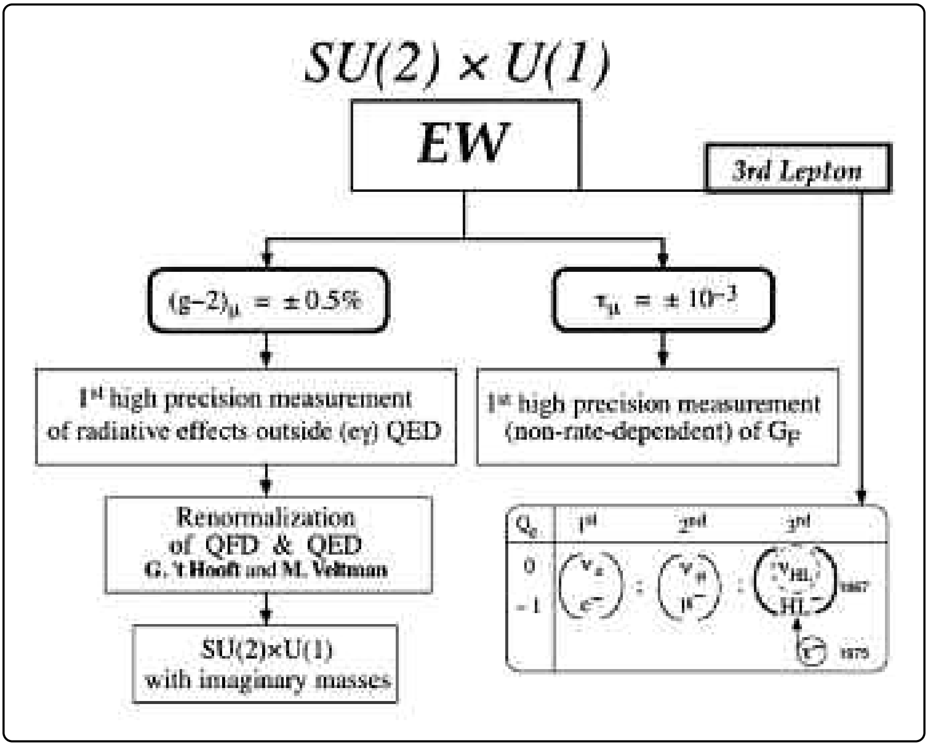


Figure 9.

POINT 2. The incredible series of events which originated with the problem of understanding the stability of matter is shown in Figures 10 and 11, together with the unexpected violation of the Symmetry Operators (C, P, T, CP) and the discovery of Matter-Antimatter Symmetry.

Mass $\neq$ Matter

$$|m_i\rangle \equiv \text{Mass} \equiv \text{Antimass} \equiv |\overline{m}_i\rangle$$

$i \equiv 1$ (Intrinsic); $i \equiv 2$ (Confinement); $i \equiv 3$ (Binding)

$$C\,|m_i\rangle = |m_i\rangle \qquad \boxed{***}$$

$$i \equiv 1, 2, 3$$

$$|m_i\, Q_i\rangle \equiv \text{Matter} \neq \text{Antimatter} \equiv |m_i\, \overline{Q}_i\rangle$$

$Q_i \equiv$ Flavour Charges

$j \equiv (u \;\; d \;\; c \;\; s \;\; t \;\; b) \equiv (1, 2, 3, 4, 5, 6)$

$(\nu_e \;\; e^- \;\; \nu_\mu \;\; \mu^- \;\; \nu_{HL} \;\; HL^-) \equiv (7, 8, 9, 10, 11, 12)$

$HL^- \downarrow \tau^-$

$$C\,|m_i\, Q_j\rangle = |m_i\, \overline{Q}_j\rangle \qquad \boxed{***}$$

$$i \equiv 1, 2, 3\,; \;\; J \equiv 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.$$

Figure 10.

There are in fact seven decades of developments which started from the antielectron and C-invariance and brought us to the discovery of nuclear antimatter and to the unification of all gauge forces with a series of unexpected discoveries, reported in Figure 11.

THE INCREDIBLE STORY
TO DISENTANGLE THE ORIGIN OF THE STABILITY OF MATTER
SEVEN DECADES FROM THE ANTIELECTRON TO ANTIMATTER
AND THE UNIFICATION OF ALL GAUGE FORCES

- **The validity of C invariance from 1927 to 1957.**

After the discovery by Thomson in 1897 of the first example of an elementary particle, the Electron, it took the genius of Dirac to theoretically discover the Antielectron thirty years after Thomson.

1927 → Dirac equation [1]; the existence of the antielectron is, soon after, theoretically predicted. Only a few years were needed, after Dirac's theoretical discovery, to experimentally confirm (Anderson, Blackett and Occhialini [2]) the existence of the Dirac antielectron.

1930-1957 → Discovery of the C operator (charge conjugation) H. Weyl and P.A.M. Dirac [3]; discovery of the P Symmetry Operator [E.P. Wigner, G.C. Wick and A.S. Wightman [4, 5]]; discovery of the T operator (time reversal) [E.P. Wigner, J. Schwinger and J.S. Bell [6, 7, 8, 9]]; discovery of the CPT Symmetry Operator from RQFT (1955-57) [10].

1927-1957 → Validity of C invariance: e^+ [2]; $\bar{p}$ [11]; $\bar{n}$ [12]; $K_2^0 \to 3\pi$ [13] but see LOY [14].

- **The new era starts: C ≠ ; P ≠ ; CP ≠ (*).**

1956 → Lee & Yang P ≠ ; C ≠ [15].

1957 → Before the experimental discovery of P ≠ & C ≠, Lee, Oehme, Yang (LOY) [14] point out that the existence of the second neutral K-meson, $K_2^0 \to 3\pi$, is proof neither of C invariance nor of CP invariance. Flavour antiflavour mixing does not imply CP invariance.

1957 → C.S. Wu et al. P ≠ ; C ≠ [16]; CP ok [17].

1964 → $K_2^0 \to 2\pi = K_L$: CP ≠ [18].

1947-1967 → QED divergences & Landau poles.

1950-1970 → The crisis of RQFT & the triumph of S-matrix theory (i.e. the negation of RQFT).

1965 → Nuclear antimatter is (experimentally) discovered [19]. See also [20].

1968 → The discovery [21] at SLAC of Scaling (free quarks inside a nucleon at very high q^2) but in violent (pp) collisions no free quarks at the ISR are experimentally found [22]. Theorists consider Scaling as being evidence for RQFT not to be able to describe the Physics of Strong Interactions. The only exception is G. 't Hooft who discovers in 1971 that the β-function has negative sign for non-Abelian theories [23].

1971-1973 → β = − ; 't Hooft, Gross & Wilczek. The discovery of **non-Abelian** gauge theories. Asymptotic freedom in the interaction between quarks and gluons [23].

1974 → All gauge couplings $\alpha_1\ \alpha_2\ \alpha_3$ run with q^2 but they do not converge towards a unique point.

1979 → A.P. & A.Z. point out that the new degree of freedom due to SUSY allows the three couplings $\alpha_1\ \alpha_2\ \alpha_3$, to converge towards a unique point [24].

1980 → QCD has a "hidden" side: the multitude of final states for each pair of interacting particles: (e^+e^- ; $p\bar{p}$; πp; Kp; νp; pp; etc.)
The introduction of the Effective Energy allows to discover the Universality properties [25] in the multihadronic final states.

1992 → All gauge couplings converge towards a unique point at the gauge unification energy: $E_{GU} \simeq 10^{16}$ GeV with $\alpha_{GU} \simeq 1/24$ [26, 27].

1994 → The Gap [28] between E_{GU} & the String Unification Energy: $E_{SU} \simeq E_{Planck}$.

1995 → CPT loses its foundations at the Planck scale (T.D. Lee) [29].

1995-1999 → No CPT theorem from M-theory (B. Greene) [30].

1995-2000 → A.Z. points out the need for new experiments to establish if matter-antimatter symmetry or asymmetry are at work.

(*) The symbol ≠ stands for 'Symmetry Breakdown'.

Figure 11.

Let me say a few words to illustrate this group of unexpected events.

Since the dawn of civilization, it was believed that the stability of matter was due to its weight: the heavier an object, the better for its stability.

The Greeks were convinced that the stability of the columns for their splendid temples was due to the 'heaviness' of the columns.

The first unexpected event started with the discovery by Einstein that mass and energy could transform each other.

> When (1905) Einstein discovered that
>
> $$mc^2 = E$$
>
> he could not sleep at night.
>
> (*Peter G. Bergmann testimony*)

If matter can transform into energy, the world becomes unstable and we could not exist.

Fortunately (for Einstein) in 1897 J.J. Thomson had discovered the 'electron'. Since its charge is opposite to that of the proton, the electric charge is enough to guarantee the stability of matter.

In fact, charge conservation forbids the transition,

$$p \rightarrow e^- \gamma ; \text{ thus } p \nrightarrow e^- \gamma . \tag{1}$$

Einstein stopped being worried and could relax; the world is stable, despite

$$E = mc^2 .$$

The only detail is that 'm' must be understood as being the 'mass' of an object, not its matter. Thus the basic distinction shown on top of Figure 10. Our Greek ancestors could not believe in this sequence of unexpected events.

And this is not all.

Another totally unexpected event had to occur a quarter of a century later, when (1930) Dirac realized that the evolution of the electron (the same particle which is in the atoms of our body) in Space-Time (with the property discovered by Lorentz in his investigation of the invariance of the Maxwell equations, i.e. they cannot both be 'real', one of them – either

Space or Time – must be 'imaginary') brought him to discover an incredibly unexpected fact: that the antielectron must exist; i.e. a very light particle with the same charge as that of the proton.

If this is the case, then the reaction

$$p \rightarrow e^{+} + \gamma \tag{2}$$

can take place and the stability of matter is again gone. The world becomes unstable.

Thus Stueckelberg invented another unexpected deviation from simplicity: the quantum number 'baryonic charge' which (according to Stueckelberg) must be conserved in nature.

The reaction (2) is forbidden by this new conservation law:

$$p \not\rightarrow e^{+} + \gamma \tag{3}$$

Life would have been far simpler if the Greeks were right in their way of explaining the stability of matter.

Despite the complications described above, the story of the stability of matter does not end with equation (3). Another unexpected discovery was needed. This came with the understanding that the word 'charge' corresponds to two totally different and fundamental physical properties.

There are in fact two types of 'charge'.

One is called 'gauge charge', and this is responsible for the existence of a Fundamental Force of Nature. There are in fact three *gauge charges*, the 'electromagnetic', the Subnuclear 'weak' and the Subnuclear 'strong'. (For simplicity we will ignore the gravitational force). The corresponding forces are described by the Gauge Symmetry Groups U(1) (for the electromagnetic forces), SU(2) (for the Subnuclear weak forces) and SU(3) (for the Subnuclear strong forces).

There is a further unexpected discovery: the two Gauge Symmetry Groups U(1) and SU(2) are mixed and it is their *mixing* which produces the effective forces which allow our TV and radio plus all electromagnetic instruments to work and the Stars to function, thanks to the effective weak force which controls very well their level of burning.

In fact, it is the strength (called Fermi coupling) of the weak forces (also called Fermi forces) which controls the amount of protons which transforms into neutrons (plus a neutrino and a positive electron) every second in a Star, thus allowing the 'nuclear fuel' (the neutrons) for the Stars to exist.

All we have said refers to the 'gauge charges', which are responsible for the existence of the fundamental forces. But the stability of matter has to be there and, contrary to what Einstein thought in 1905, it is not guaranteed by the existence of the electric charge: this is in fact a 'gauge charge'.

In order to have the stability of matter, we need a totally unexpected type of charge, called *flavour charge*. The number of these charges is again unexpectedly 12, as illustrated in Figure 10. The incredible story is that these charges are four times more than the 'gauge charges'; so far nobody knows why this is so.

The answer could come from the Superworld.

The conclusion of the long way to understand the origin of the stability of matter is that many unexpected discoveries were needed, as reported in Figure 11. From the Greeks who associated 'stability' of matter with 'heaviness' to our present understanding, the number of unexpected events is really impressive.

Let us skip the points 3 and 4 and move to 5, the Physics of Instantons in QCD.

POINT 5. The mixing in the pseudoscalar and in the vector mesons: the Physics of Instantons (Figure 12).

In the Physics of Mesons the totally unexpected result was the difference existing between the two mesonic mixing angles, pseudoscalar and vector:

$$\theta_{PS} \neq \theta_V .$$

They should both be zero if $SU(3)_{uds}$ was a good Symmetry. The existence of Instantons was not known. They came after the unexpected discovery that $\theta_{PS} \neq \theta_V$. (See Figure 12, page 287)

POINT 6. Newton discovered that Light is the sum of different Colours. The modern mathematical structure which describes light is Quantum ElectroDynamics: QED. In the Subnuclear world of *quarks and gluons* the mathematical structure which describes these forces is called Quantum ChromoDynamics. The interaction between *quarks and gluons* produce *Jets* made of many *pions*

$$pp \rightarrow \pi + X .$$

The energy spectrum of these *pions* is shown in Figure 13.

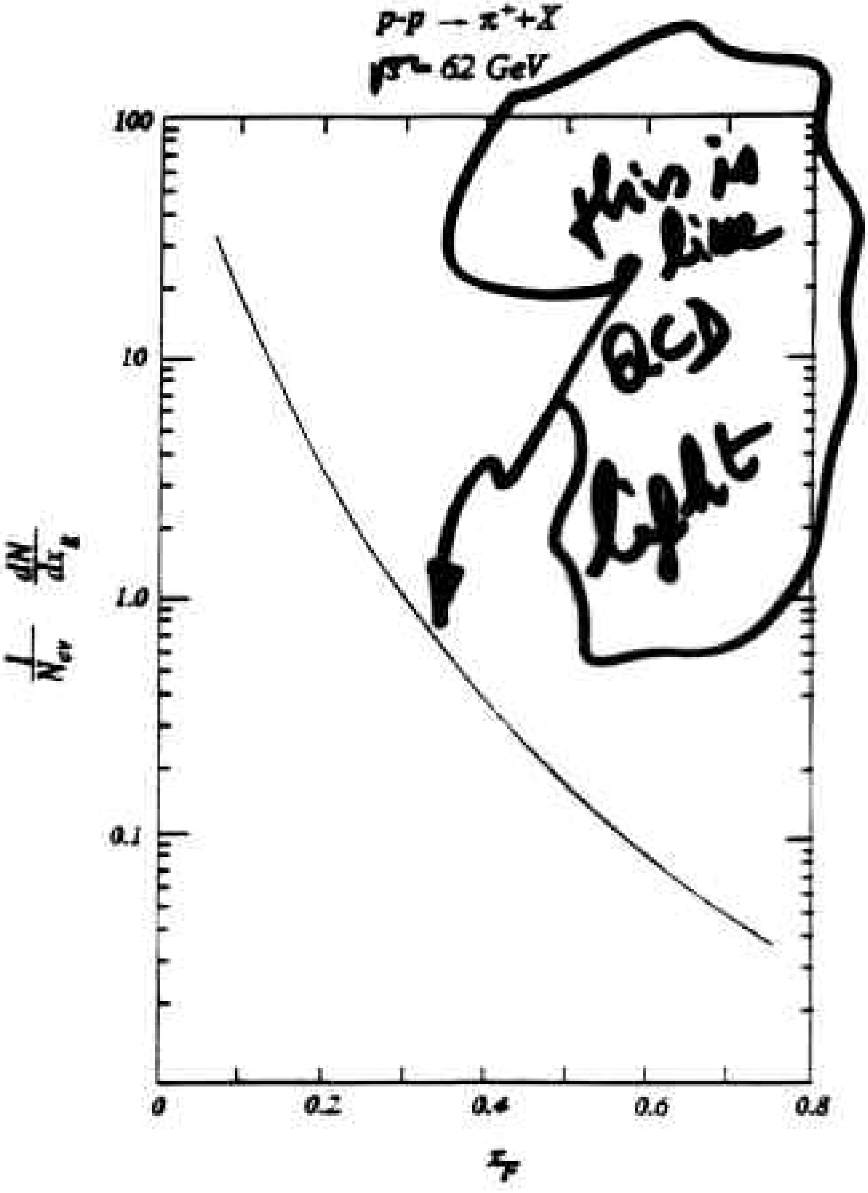

p-p → π⁺+X
√s = 62 GeV
100
10
1.0
0.1
0
0.2
0.4
0.6
0.8
$\frac{1}{N_{ev}} \frac{dN}{dx_F}$
x_F
this is
lime
QCD
light

Figure 13.

This is what Gribov defined: *The QCD light*. In fact this unique 'line' is the sum of all 'lines' shown in Figure 14 (see page 288). This is the discovery of the Effective Energy, once again a totally unexpected quantity.

The non-Abelian nature of the Interaction describing quarks, gluons and the Effective Energy with the set of unexpected discoveries is illustrated in Figure 15.

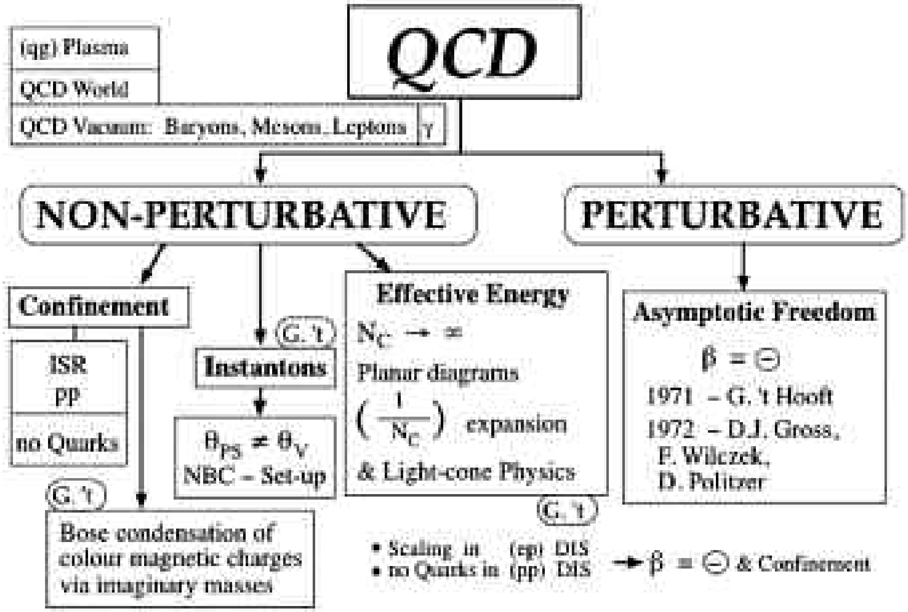


Figure 15.

POINT 7. The Unification of all Forces and the Supersymmetry threshold with its problems are reported in Figures 16 and 17 (see pages 289-90) respectively. This Figure illustrates the EGM effect in bringing down by a factor 700 the threshold for the production of the lightest superparticle.

The mathematical formalism which has been used to obtain the results shown in these Figures is a system of three differential non-linear equations coupled via the gauge couplings

α_i , α_j (with i = 1,2,3; and J = 1,2,3 but $i \neq j$),

as shown in Figure 18.

THE UNIFICATION OF ALL FUNDAMENTAL FORCES

$$\mu \frac{d\alpha_i}{d\mu} = \frac{b_i}{2\pi} \alpha_i^2 + \sum_j \frac{b_{ij}}{8\pi^2} \alpha_i \alpha_j$$

Figure 18.

For more than ten years (from 1979 to 1991), no one had realized that the energy threshold for the existence of the Superworld was strongly dependent on the 'running' of the masses.

This is now called: the EGM effect (from the initials of Evolution of Gaugino Masses).

To compute the energy threshold using only the 'running' of the gauge couplings (α_1, α_2, α_3) corresponds to neglecting nearly three orders of magnitude in the energy threshold for the discovery of the first particle (the lightest) of the Superworld [30], as illustrated in Figure 17.

Let me now illustrate in Figure 19 (see page 291) the Platonic Grand Unification in the Convergence of the three Gauge Couplings $\alpha_1, \alpha_2, \alpha_3$.

The 'Platonic' Simplicity would indicate the series of points making up the straight line as the ideally simple solution. The real solution is the sequence of points which totally deviate from the straight line.

This is just an example of comparison between the 'Platonic' Simplicity and the 'real world', with the unexpected events needed to deal with the Grand Unification (see Figure 16, page 289).

Talking about Supersymmetry, there is another important step: how we go from pure Theoretical Speculations to Phenomenology.

The proof is given in Figure 20 where it is shown how many important properties of the physics to be described have been neglected by some authors (AdBF) whose claim was to 'predict' the energy scale at which Supersymmetry is broken.

In order to attempt to give such a prediction, there are at least five 'details' to be taken into account, as reported in the last five columns of Figure 20.

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩
Authors	Input data	Errors	EC	M_{SUSY}	CC	UC	ΔT_L	M_X	ΔT_H	EGM
ACPZ [47, 49–54]	WA	$\pm 2\sigma$	all possible solutions (26)	Yes	physical	Yes	Yes	Yes	Yes	Yes

Authors	Input data	Errors	EC	M_{SUSY}	CC	UC	ΔT_L	M_X	ΔT_H	EGM
AdBF [55]	only one experiment	$\pm 1\sigma$	only one solution	Yes	Geometrical	No	No	No	No	No

① WA = World Average
② Errors = Uncertainty taken from all data (World Average) or from a single experiment
③ EC = Evolution of Couplings
④ M_{SUSY} = Mass Scale assumed to represent the Supersymmetry Scale breaking
⑤ CC = Convergence of Couplings
⑥ UC = Unification of Couplings above E_{GUT}
⑦ ΔT_L = Low Energy threshold
⑧ M_X = Mass Scale at the breaking of the Grand Unified Theory to the SU(3)xSU(2)xU(1)
⑨ ΔT_H = High Energy threshold
⑩ EGM = Evolution of Gaugino Masses

Figure 18.

4. Conclusions

In the field of Subnuclear Physics, the totally unexpected discoveries date back to the beginning of Galilean Science. We have listed those of the 20th Century ending up with present-day frontiers in Subnuclear Physics.

Question. *What about other fields?* One which is very intensive in a number of discoveries is the field of condensed matter.

Let me quote Tony Leggett (University of Illinois, Urbana – Champaign, USA), Nobel 2003 for 'Superfluidity': 'It is relatively rare in Condensed-Matter Physics to predict discoveries, it is a field where you fall over them by accident'.

REFERENCES

[1] P.A.M. Dirac, 'The Quantum Theory of the Electron', *Proc. Roy. Soc.* (London) A117, 610 (1928); 'The Quantum Theory of the Electron, Part II', *Proc. Roy. Soc.* (London) A118, 351 (1928).

[2] C.D. Anderson, 'The Positive Electron', *Phys. Rev.* 43, 491 (1933); 'Some Photographs of the Tracks of Penetrating Radiation', P.M.S. Blackett and G.P.S. Occhialini, *Proc. Roy. Soc.* A139, 699 (1933).

[3] H. Weyl, *Gruppentheorie und Quantenmechanik*, 2nd ed., 234 (1931).

[4] E.P. Wigner, 'Unitary Representations of the Inhomogeneous Lorentz Group', *Ann. Math.*, 40, 149 (1939).

[5] G.C. Wick, E.P. Wigner, and A.S. Wightman, 'Intrinsic Parity of Elementary Particles', *Phys. Rev.* 88, 101 (1952).

[6] E.P. Wigner, 'Über die Operation der Zeitumkehr in der Quantenmechanik', *Gött. Nach.* 546-559 (1931). Here for the first time an anti-unitary symmetry appears.

[7] E.P. Wigner, *Ann. Math.* 40, 149 (1939).

[8] J. Schwinger, *Phys. Rev.* 82, 914 (1951).

[9] J.S. Bell, 'Time Reversal in Field Theory', *Proc. Roy. Soc.* (London) A231, 479-495 (1955).

[10] To the best of my knowledge, the CPT theorem was first proved by W. Pauli in his article 'Exclusion Principle, Lorentz Group and Reflection of Space-Time and Charge', in *Niels Bohr and the Development of Physics* [Pergamon Press, London, page 30 (1955)], which in turn is an extension of the work of J. Schwinger [*Phys. Rev.* 82, 914 (1951); 'The Theory of Quantized Fields. II.', *Phys. Rev.* 91,

713 (1953); 'The Theory of Quantized Fields. III.', *Phys. Rev.* 91, 728 (1953); 'The Theory of Quantized Fields. VI.', *Phys. Rev.* 94, 1362 (1954)] and G. Lüders, 'On the Equivalence of Invariance under Time Reversal and under Particle-Anti-particle Conjugation for Relativistic Field Theories' [*Dansk. Mat. Fys. Medd.* 28, 5 (1954)], which referred to an unpublished remark by B. Zumino. The final contribution to the CPT theorem was given by R. Jost, in 'Eine Bemerkung zum CPT Theorem' [*Helv. Phys. Acta* 30, 409 (1957)], who showed that a weaker condition, called 'weak local commutativity' was sufficient for the validity of the CPT theorem.

[11] 'Observation of Antiprotons', O. Chamberlain, E. Segrè, C. Wiegand, and T. Ypsilantis, *Physical Review* 100, 947 (1955).

[12] 'Anti-Neutrons Produced from Anti-Protons in Charge Exchange Collisions', B. Cork, G.R. Lambertson, O. Piccioni, W.A. Wenzel, *Physical Review* 104, 1193 (1957).

[13] 'Observation of Long-Lived Neutral V Particles', K. Lande, E.T. Booth, J. Impeduglia, L.M. Lederman, and W. Chinowski, *Physical Review* 103, 1901 (1956).

[14] 'Remarks on Possible Noninvariance under Time Reversal and Charge Conjugation', T.D. Lee, R. Oehme, and C.N. Yang, *Physical Review* 106, 340 (1957).

[15] 'Question of Parity Conservation in Weak Interactions', T.D. Lee and C.N. Yang, *Phys. Rev.* 104, 254 (1956).

[16] 'Experimental Test of Parity Conservation in Beta Decay', C.S. Wu, E. Ambler, R.W. Hayward, D.D. Hoppes, *Phys. Rev.* 105, 1413 (1957); 'Observation of the Failure of Conservation of Parity and Charge Conjugation in Meson Decays: The Magnetic Moment of the Free Muon', R. Garwin, L. Lederman, and M. Weinrich, *Phys. Rev.* 105, 1415 (1957); 'Nuclear Emulsion Evidence for Parity Non-Conservation in the Decay Chain $\pi^+\mu^+e^+$', J.J. Friedman and V.L. Telegdi, *Phys. Rev.* 105, 1681 (1957).

[17] 'On the Conservation Laws for Weak Interactions', L.D. Landau, *Zh. Éksp. Teor. Fiz.* 32, 405 (1957).

[18] 'Evidence for the 2π Decay of the K_2^0 Meson', J. Christenson, J.W. Cronin, V.L. Fitch, and R. Turlay, *Physical Review Letters* 113, 138 (1964).

[19] 'Experimental Observation of Antideuteron Production', T. Massam, Th. Muller, B. Righini, M. Schneegans, and A. Zichichi, *Nuovo Cimento* 39, 10 (1965).

[20] 'The Discovery of Nuclear Antimatter', L. Maiani and R.A. Ricci (eds.), Conference Proceedings 53, Italian Physical Society, Bologna, Italy (1995); see also A. Zichichi in *Subnuclear Physics – The first fifty years* (2000).
[20] 'The Discovery of Nuclear Antimatter', L. Maiani and R.A. Ricci (eds.), Conference Proceedings 53, Italian Physical Society, Bologna, Italy (1995); see also A. Zichichi in *Subnuclear Physics – The First Fifty Years*, O. Barnabei, P. Pupillo and F. Roversi Monaco (eds.), a joint publication by University and Academy of Sciences of Bologna, Italy (1998); World Scientific Series in 20th Century Physics, Vol. 24 (2000).
[21] The first report on 'scaling' was presented by J.I. Friedman at the 14th International Conference on High Energy Physics in Vienna, 28 August-5 September 1968. The report was presented as paper n. 563 but not published in the Conference Proceedings. It was published as a SLAC preprint. The SLAC data on scaling were included in the Panofsky general report to the Conference where he says '... the apparent success of the parametrization of the cross-sections in the variable ν/q^2 in addition to the large cross-section itself is at least indicative that point-like interactions are becoming involved'. 'Low q^2 Electrodynamics, Elastic and Inelastic Electron (and Muon) Scattering', W.K.H. Panofsky in Proceedings of 14th International Conference on High Energy Physics in Vienna 1968, J. Prentki and J. Steinberger (eds.), page 23, published by CERN (1968). The following physicists participated in the inelastic electron scattering experiments: W.B. Atwood, E. Bloom, A. Bodek, M. Breidenbach, G. Buschhorn, R. Cottrell, D. Coward, H. DeStaebler, R. Ditzler, J. Drees, J. Elias, G. Hartmann, C. Jordan, M. Mestayer, G. Miller, L. Mo, H. Piel, J. Poucher, C. Prescott, M. Riordan, L. Rochester, D. Sherden, M. Sogard, S. Stein, D. Trines, and R. Verdier. For additional acknowledgements see J.I. Friedman, H.W. Kendall and R.E. Taylor, 'Deep Inelastic Scattering: Acknowledgements', Les Prix Nobel 1990, (Almqvist and Wiksell, Stockholm/Uppsala 1991), also *Rev. Mod. Phys*. 63, 629 (1991). For a detailed reconstruction of the events see J.I. Friedman 'Deep Inelastic Scattering Evidence for the Reality of Quarks' in History of Original Ideas and Basic Discoveries in Particle Physics, H.B. Newman and T. Ypsilantis (eds.), Plenum Press, New York and London, 725 (1994).
[22] *Quark Search at the ISR*, T. Massam and A. Zichichi, CERN preprint, June 1968; 'Search for Fractionally Charged Particles Produced in

Proton-Proton Collisions at the Highest ISR Energy', M. Basile, G. Cara Romeo, L. Cifarelli, P. Giusti, T. Massam, F. Palmonari, G. Valenti and A. Zichichi, Nuovo Cimento 40A, 41 (1977); and 'A Search for quarks in the CERN SPS Neutrino Beam', M. Basile, G. Cara Romeo, L. Cifarelli, A. Contin, G. D'Alì, P. Giusti, T. Massam, F. Palmonari, G. Sartorelli, G. Valenti and A. Zichichi, Nuovo Cimento 45A, 281 (1978).

[23] A. Zichichi in *Subnuclear Physics – The First Fifty Years*, O. Barnabei, P. Pupillo and F. Roversi Monaco (eds.), a joint publication by University and Academy of Sciences of Bologna, Italy (1998); World Scientific Series in 20th Century Physics, Vol. 24 (2000).

[24] 'New Developments in Elementary Particle Physics', A. Zichichi, *Rivista del Nuovo Cimento* 2, n. 14, 1 (1979). The statement on page 2 of this paper, 'Unification of all forces needs first a Supersymmetry. This can be broken later, thus generating the sequence of the various forces of nature as we observe them', was based on a work by A. Petermann and A. Zichichi in which the renormalization group running of the couplings using supersymmetry was studied with the result that the convergence of the three couplings improved. This work was not published, but perhaps known to a few. The statement quoted is the first instance in which it was pointed out that supersymmetry might play an important role in the convergence of the gauge couplings. In fact, the convergence of three straight lines ($\alpha_1^{-1}\alpha_2^{-1}\alpha_3^{-1}$) with a change in slope is guaranteed by the Euclidean geometry, as long as the point where the slope changes is tuned appropriately. What is incorrect about the convergence of the couplings is that, with the initial conditions given by the LEP results, the change in slope needs to be at $M_{SUSY} \sim 1$ TeV as claimed by some authors not aware in 1991 of what was known in 1979 to A. Petermann and A. Zichichi.

[25] V.N. Gribov, G. 't Hooft, G. Veneziano and V.F. Weisskopf, *The Creation of Quantum ChromoDynamics and the Effective Energy*, L.N. Lipatov (ed.), a joint publication by the University and the Academy of Sciences of Bologna, Italy (1998); World Scientific Series in 20th Century Physics, Vol. 25 (2000).

[26] 'The Effective Experimental Constraints on M_{SUSY} and M_{GUT}', F. Anselmo, L. Cifarelli, A. Petermann and A. Zichichi, *Nuovo Cimento* 104A, 1817 (1991).

[27] 'The Simultaneous Evolution of Masses and Couplings: Consequences on Supersymmetry Spectra and Thresholds', F. Anselmo, L. Cifarelli,

A. Petermann and A. Zichichi, *Nuovo Cimento* 105A, 1179 (1992).
[28] 'A Study of the Various Approaches to M_{GUT} and α_{GUT}', F. Anselmo, L. Cifarelli and A. Zichichi, *Nuovo Cimento* 105A, 1335 (1992).
[29] 'Are Matter and Antimatter Symmetric?', T.D. Lee, in *Proceedings of the Symposium to celebrate the 30th anniversary of the Discovery of Nuclear Antimatter*, L. Maiani and R.A. Ricci (eds.), Conference Proceedings 53, p. 1, Italian Physical Society, Bologna, Italy (1995).
[30] 'String Theory: the Basic Ideas', B. Greene, Erice Lectures – Discussion 1999 in *Basics and Highlights in Fundamental Physics*, A. Zichichi (ed.), World Scientific (2001).
[31] 'Search for Supersymmetric Particles using Acoplanar Charged Particle Pairs from Z^0 decays', ALEPH Collab., D. Decamp *et al.*, *Phys. Lett*. B236, 86 (1990).
[32] 'Search for Neutral Higgs Bosons from Supersymmetry in Z decays' ALEPH Collab., D. Decamp *et al.*, *Phys. Lett*. B237, 291 (1990).
[33] 'Search for Neutralino Production in Z decays', ALEPH Collab., D. Decamp *et al.*, *Phys. Lett*. B244, 541 (1990).
[34] 'Search for the Neutral Higgs Bosons of the MSSM and other two Doublet Models', ALEPH Collab., D. Decamp *et al.*, *Phys. Lett*. B265, 475 (1991).
[35] 'Search for Heavy Charged Scalars in Z^0 decays', DELPHI Collab., P. Abreu *et al.*, *Phys. Lett*. B241, 449 (1990).
[36] 'Search for Pair Production of Neutral Higgs Bosons in Z^0 decays', DELPHI Collab., P. Abreu *et al.*, *Phys. Lett*. B245, 276 (1990).
[37] 'Search for Scalar Quarks in Z^0 decays', DELPHI Collab., P. Abreu *et al.*, *Phys. Lett*. B247, 148 (1990).
[38] 'A Search for Sleptons and Gauginos in Z^0 Decays', DELPHI Collab., P. Abreu *et al.*, *Phys. Lett*. B247, 157 (1990).
[39] 'Mass Limits for Scalar Muons, Scalar Electrons and Winos from e^+e^- Collisions near S**(1/2) = 91–GeV', L3 Collab., B. Adeva *et al.*, *Phys. Lett*. B233, 530 (1989).
[40] 'Search for the Neutral Higgs Bosons of the Minimal Supersymmetric Standard Model from Z^0 Decays', L3 Collab., B. Adeva *et al.*, *Phys. Lett*. B251, 311 (1990).
[41] 'Search for the Charged Higgs Boson in Z^0 decay', L3 Collab., B. Adeva *et al.*, *Phys. Lett*. B252, 511 (1990).
[42] 'A Search for Acoplanar Pairs of Leptons or Jets in Z^0 decays: Mass Limits on Supersymmetric Particles', OPAL Collab., M.Z. Akrawy *et al.*, *Phys. Lett*. B240, 261 (1990).

[43] 'A Search for Technipions and Charged Higgs Bosons at LEP', OPAL Collab., M.Z. Akrawy *et al.*, *Phys. Lett*. B242, 299 (1990).

[44] 'A Direct Search for Neutralino Production at LEP', OPAL Collab., M.Z. Akrawy *et al.*, *Phys. Lett.* B248, 211 (1990); P.D. Acton *et al.*, preprint CERN-PPE/91-115, 22 July 1991.

[45] 'Searches for Supersymmetric Particles Produced in Z Boson decay', MARK II Collab., T. Barklow *et al.*, *Phys. Rev. Lett.* 64, 2984 (1990).

[46] 'Searches for New Particles at LEP', M. Davier, LP-HEP 91 Conference, Geneva, CH, Preprint LAL 91-48, December 1991.

[47] 'The Evolution of Gaugino Masses and the SUSY Threshold', F. Anselmo, L. Cifarelli, A. Peterman and A. Zichichi, *Nuovo Cimento* 105A, 581 (1992).

[48] 'A Detailed Comparison of LEP Data with the Predictions of the Minimal Supersymmetric SU(5) GUT', J.R. Ellis, S. Kelley, D.V. Nanopoulos, preprint CERN-TH/6140-91, *Nucl. Phys*. B373, 55 (1992).

[49] 'The Effective Experimental Constraints on M_{SUSY} and M_{GUT}', F. Anselmo, L. Cifarelli, A. Peterman and A. Zichichi, *Nuovo Cimento* 104A, 1817 (1991).

[50] 'The Convergence of the Gauge Couplings at E_{GUT} and above: Consequences for $\alpha_3(M_Z)$ and SUSY Breaking', F. Anselmo, L. Cifarelli, A. Peterman and A. Zichichi, *Nuovo Cimento* 105A, 1025 (1992).

[51] 'The Simultaneous Evolution of Masses and Couplings: Consequences on Supersymmetry Spectra and Thresholds', F. Anselmo, L. Cifarelli, A. Peterman and A. Zichichi, *Nuovo Cimento* 105A, 1179 (1992).

[52] 'Analytic Study of the Supersymmetry-Breaking Scale at Two Loops', F. Anselmo, L. Cifarelli, A. Peterman and A. Zichichi, *Nuovo Cimento* 105A, 1201 (1992).

[53] 'A Study of the Various Approaches to M_{GUT} and α_{GUT}', F. Anselmo, L. Cifarelli and A. Zichichi, *Nuovo Cimento* 105A, 1335 (1992).

[54] 'A χ^2-Test to Study the α_1, α_2, α_3 Convergence for High-Precision LEP Data, Having in Mind the SUSY Threshold', F. Anselmo, L. Cifarelli and A. Zichichi, *Nuovo Cimento* 105A, 1357 (1992).

[55] U. Amaldi, W. de Boer and H. Furstenau, *Phys. Lett.* B260, 447 (1991).

DISCOVERING THE WORLD STRUCTURE AS A GOAL OF PHYSICS

MICHAEL HELLER

SUMMARY

The structuralist view on mathematics, claiming that mathematics is 'about structures and their morphology', is well founded in both mathematical practice and metatheoretical investigation, although it is difficult to formulate in a rigorous manner. Assuming the structuralist view on mathematics, we look for its consequences for interpreting physical theories.

Empirical successes of physical theories support their realistic interpretation, but discontinuities in the world image, often caused by replacing a successful physical theory by another even more successful theory, support their anti-realistic interpretation. However, if we remember that the mathematical structure of the old theory can be obtained, as a limiting case, from the mathematical structure of the new theory, we recover a continuity at the level of structures (even if there is no continuity at the level of respective world images). This supports the view that physical theories are 'about the structure of the world' or, more precisely, that the method of physics grasps only the world structure. From the philosophical point of view, two possibilities can be envisaged: either (1) structures discovered by physical theories are structures of something, i.e., there is a 'structured stuff', but this stuff is transparent for the physical method (epistemic structuralism), or (2) such a stuff does not exist (ontic structuralism). To decide between these two possibilities one should go beyond the analysis of the physical method and appeal to metaphysical or logical reasons.

In the case, when one physical theory admits more than one mathematical formulation (which is not infrequent situation in physics), we should assume that these formulations are but various representations of an abstract structure, and that it is this abstract structure that reflects or approximates the Structure of the World.

1. Introduction

The line of reasoning adopted in this paper is the following:

– Mathematics is a science of structures, i.e., the world discovered (or constructed?) by mathematicians consists of structures and relations between structures.

– Physics employs mathematical structures to model the world.

– Therefore, the world as discovered (or constructed?) by physicists consists of mathematical structures interpreted as structures of the world.

As we can see, the above strategy is based on methodology: by analysing the method of both mathematics and physics we try to gain some important information about the world. And the information is that the world, as discovered (or constructed) by physics, is purely structural. Only its structural aspects are visible for the mathematical-experimental method of physics. A material stuff supporting relations, that constitute the structure, if it exists, is transparent for this method.

This is not a new philosophy of physics. It is known for a long time that there are interactions between physical objects rather than the objects themselves that are accessible for the method of physics. And the word 'interactions' is just another name for 'relations constituting the structure'. However, in the last decades, this philosophy of physics was subject to a detailed scrutiny, and became a subject-matter of a long dispute, out of which new and illuminating arguments have emerged.

The above presented 'line of reasoning' determines the organization of my material. First, I focus on the 'mathematical structuralism', then I discuss its use in physics to finally draw some conclusions concerning what could be called the structuralist ontology of the world.

In pursuing this plan, I base my presentation on the above-mentioned debate which is still going on in the foundations of mathematics and in philosophy of physics.

2. Structuralist Approach to Mathematics

It would be difficult to find a mathematician who, for the first time, used the word 'structure' to denote the subject-matter of mathematical inquiry. Nowadays, the saying that mathematics investigates structure is a kind of common wisdom among mathematicians. Any deeper study of mathematics, especially of geometry, creates a persistent impression of an architectonic totality, the elements of which acquire their meaning only through their relationship with other elements and the underlying plan. No wonder that structuralist ideas had appeared long before they became a part of an official current in the philosophy of mathematics. The concept of structure (understood in a more technical way) invaded first the abstract algebra, and from algebra it proliferated – mainly owing to the works of Nicolai Bourbaki – to almost the whole of mathematics. Some structuralist view could be traced back to the writings of Hilbert, Bernays and Quine, but it is often Michael Resnik who is credited with initiating the 'structuralist movement'. In his paper 'Mathematics as Science of Patterns: Ontology and Reference',[1] considered by many as a kind of structuralist manifesto, Resnik declared that in mathematics we are never dealing with objects equipped with 'inner properties', but only with structures. The objects, studied in mathematics, are but 'structureless points' or positions within structures. Besides their place in a structure, they are devoid of any individuality and any property independent of the structure. He wrote:

> The objects of mathematics, that is, the entities which our mathematical constants and quantifiers denote, are structureless points or positions in structures. As positions in structures, they have no identity outside of a structure.[2]

The thesis, understood intuitively, that mathematics is a science of structure, does not excite any emotions, but if one wants to make it more precise, opinions start to diverge, and technical discussions replace a consensus. As far as these technical aspects of the problem are concerned two mathematical theories seem to be especially relevant: set theory and category theory.

2.1. *Mathematical Structuralism and Set Theory*

The great majority of current works in mathematics is based on the assumption that mathematics is founded on set theory. No wonder, therefore, that when we think on the structural approach to mathematics, the first question that should be answered is whether the set theoretic definition of number (which 'evidently' are objects!) could be rephrased in a structuralist manner. Paul Benacerraf, in his paper 'What Numbers Could

[1] *Nous*, 15 (1981), pp. 529-550.
[2] *Ibid.*, p. 530.

Not Be?',[3] addresses this question. He points out that, for instance, number '3' can be identified with different set theoretic objects (e.g. with {{{0}}} in agreement with Zermelo's approach or with {0, {0}, {0}} in agreement with von Neumann's approach), and any such 'object' can serve as a representative of number '3' provided it satisfies certain structural relations. The mathematician's interest does not go beyond structural properties. It is impossible to identify an 'object' independently of the role it plays in the structure. Referring to this view, Charles Parsons[4] speaks about the eliminative structuralism: all utterances on objects should be understood as utterances on structures, and one should look for mathematical procedures that would allow one to eliminate from the discourse all mention on objects by substituting for them structural terms.

Structuralist approach to the philosophy of mathematics is often identified with the Platonist approach. In fact, Resnik, in his seminal paper, claimed that it is the structuralist approach that puts the Platonist philosophy on the firm ground. In his view, the belief in objects is incompatible with the Platonist philosophy since 'no mathematical theory can do more than determine its objects up to isomorphism', and

> [t]he Platonist seems to be in the paradoxical position of claiming that a given mathematical theory is about certain things and yet [he is] unable to make any definitive statement of what these things are.[5]

However, this link between structuralism and Platonism was questioned by Steward Shapiro.[6] He argues that since 'mathematics applies to reality through the discovery of mathematical structures underlying the non-mathematical universe', the latter is ontologically prior with respect to mathematical structures. Although Shapiro's views are open to a criticism,[7] it is important to realize that mathematical structuralism is not to be identified with mathematical Platonism.

The strict definition of structure can be found in abstract algebra. Here structure is understood as a domain, possibly with some distinguished elements, on which some relations and functions are defined and

[3] In P. Benacerraf, H. Putnam (eds.), *Philosophy of Mathematics: Selected Readings*, 2nd edition (Cambridge, Cambridge University Press, 1983), pp. 272-294 (1st edition, 1965).

[4] 'The Structuralist View of Mathematical Objects', *Synthese*, 84 (1990), pp. 303-306.

[5] *Op. cit.*, p. 529.

[6] 'Mathematics and Reality', *Philosophy of Science*, 50 (1983), pp. 523-548.

[7] See, e.g., Ch. Chihara, *Constructability and Mathematical Existence* (Oxford, Clarendon Press, 1990).

satisfy suitable axioms. Examples of such structures are: group, vector space, module, linear algebra. However, if one wants to use this structure definition for philosophical purposes, one immediately meets the following difficulty: the domain entering the structure definition is a set, and relations and functions on this domain are defined in terms of Cartesian products of sets. It is, therefore, evident that this structure definition presupposes the existence of sets, and cannot serve as a tool to replace the 'set philosophy' with a structural approach. Problems of this sort make people look for a solution in category theory.

2.2. *Mathematical Structuralism and Category Theory*

One of the main motives to create category theory was to grasp the essence of structure, S. Eilenberg and S. MacLane[8] aimed at this topic not for philosophical reasons but to elaborate a tool to investigate dependences between various mathematical theories.

For precise definitions introducing the concept of category the reader should look in the suitable mathematical literature,[9] here we give only an 'intuitive description', sufficient to follow further analyses. By a *category* one should understand:

(1) a class of objects: U, V, W, Z,...

(2) a class of *morphisms*: *Mor (U, V)* for every pair (U, V) of objects,

(3) a composition of morphisms.

Morphisms could be viewed as mappings between objects, although neither objects have to be sets, nor morphisms have to be mappings between sets. Of course, objects, morphisms and the operation of composition of morphisms must satisfy suitable axioms.

Common examples of categories are: category *Ens* with any sets as objects, and any mapping between them as morphisms; category *Top* with topological spaces as objects and continuous mappings as morphisms; category *Gr* with (non-empty) group as objects, and group homomorphisms as morphisms.

An important role in category theory is played by *functors*; they could be thought of as mappings between categories that preserve some of their

[8] 'General Theory of Natural Equivalences', *Trans. Amer. Math. Soc.,* 58 (1945), pp. 231-294.

[9] A good introductory course could be found in R. Geroch, *Mathematical Physics* (Chicago-London, The University of Chicago Press, 1985).

properties. For instance, the *forgetting functor* goes from *Gr* to *Ens* in such a way that it ascribes to every group the set on which the group is defined (this functor preserves the set structure but forgets the group structure).

In category theory the stress is laid on morphisms and functors rather than on objects. Motivated by this intuitive idea, suggestions were made that category theory could provide new foundations for mathematics.[10] It is interesting to notice that objections against this view were directed not against structuralism, but rather in defense of the thesis that it is only set theory that could provide necessary tools to reach this goal.[11] However, this claim seems to be unjustified because of the existence of the topos category. It is a category that unifies in itself the notion of a 'generalized set' with the notion of a 'generalized space'.[12] There are many topoi which can be used for various purposes. For example, one can define a topos that does not require the law of the excluded middle, or a topos without the axiom of choice. Among these topoi there is one which is equivalent to the standard set theory. Because of this elasticity the theory of topoi combines in itself properties of geometry and logic. It is therefore evident that everything that can be achieved for the foundations of mathematics with set theory can also be done with category theory. Could category theory do more than set theory in creating structuralist foundations of mathematics? The situation in this respect is not clear.

The evolution of MacLane's views in this respect is significant. In his earlier work[13] he expressed a more radical view, and hoped that category theory would be able to provide strict and better foundations of mathematics than those based on set theory. Later on, when these hopes had not materialized, he had to change his standpoint. Out of the failure he made a philosophy. Mathematics has a 'Protean character', in the sense that it explains why it can have no foundations. Although category theory cannot provide strict foundations for mathematics, it has a 'foundational significance' since it organizes the whole of mathematics so that we are able to speak about mathematical structures and structures of these structures.[14]

[10] S. MacLane, *Mathematics: Form and Function* (New York, Springer, 1986).

[11] J. Mayberry, 'What is Required of a Foundations of Mathematics?', *Philosophia Mathematica*, 2 (1992), pp. 16-35.

[12] For precise definitions see, for instance: S. MacLane, I. Moerdijk, *Sheaves in Geometry and Logic: A First Introduction to Topos Theory* (New York, Springer, 1992).

[13] *Mathematics: Form and Function*, *op. cit*.

[14] S. MacLane, 'The Protean Character of Mathematics', *The Space of Mathematics*: J. Echeverra, A. Ibarra, J. Mormann, De Gruyter (eds.), (New York, 1997), pp. 3-12.

In this sense, 'mathematics has as its subject-matter structures and their morphology'.[15] To express these ideas in a more strict language people start to speak about the 'category of all categories'. Attempts to give its precise definition are involved in many technicalities, the discussion of which would lead us beyond the scope of the present paper.

Finally, let us mention that there were attempts at creating a formal theory of structure. For instance, Shapiro proposed a suitable axiomatic system.[16] However, since his system imitates that of a set theory, it is involved essentially in the same interpretative difficulties.[17]

2.3. *Tentative Conclusions*

Let us collect some tentative conclusions following from the above analysis. The structuralist interpretation of mathematics is well-founded in both everyday mathematical practice and in metatheoretical investigations. However, it is notoriously difficult to express it in a strict but adequate way. Moreover, the border-line between the 'objectivist' and 'structuralist' views is rather fuzzy, and if one wanted, in some cases, to determine this border more precisely, one would have to look at what a given mathematician is doing rather than to listen to his or her declarations.

We should distinguish the weaker and stronger versions of mathematical structuralism. Adherents of the weaker version do not negate the existence of mathematical objects but treat them as 'places' in a structure. These places are either devoid of an 'inner structure', or their structures are entirely determined by the structure in which they are substructures. Adherents of the stronger version claim that either the concept of structure does not require the existence of objects at all (ontological structuralism), or that if objects do exist, they are not cognizable. It is clear that the latter distinction have strong metaphysical bearing, and goes beyond the foundations of mathematics.

In the following, we shall work with the assumption that mathematics is 'about structures and their morphology', and look for the consequences

[15] E. Landry, 'Category Theory: The Language of Mathematics', scistud.umkc.edn/psa 98/papers/.

[16] See, S. Shapiro, *Philosophy of Mathematics. Structure and Ontology* (New York-Oxford, Oxford University Press, 1997).

[17] For a detailed analysis see: K. Wójtowicz, *Spór o istnienie w Matematyce (Dispute on the Existence in Mathematics)*, in Polish (Warszawa, Semper, 2003), chapter 10.C.

of this assumption as far as the nature of physics is concerned. We shall, in principle, understand this assumption in the weaker sense, but we shall also occasionally try to see the implications of its stronger forms.

3. The Best of Both Worlds

If mathematics is a science of structures, and physics' main tool in investigating the world is by constructing its mathematical models, the question arises in which sense the structuralist view on mathematics transfers into our knowledge of the physical world. It is a happy circumstance that in order to elucidate this question, we can make use of a recent discussion in the philosophy of physics concerning these matters. Although this discussion was focused on a link between the structuralist approach to physics and its realist or antirealist interpretations, it also significantly contributed to a better understanding of the structuralist standpoint itself.

Remarks made by Jeremy Butterfield, a philosopher of science, and Chris Isham, a theoretical physicist, in their recent paper on quantum gravity[18] can serve as a good introduction to this discussion. When warning against dangers of a far-fetched realism in interpreting physics, they claim that every major innovation in physics changes its very subject-matter. In their opinion, this follows from the fact that physics aims at giving a complete description of the world. Both the common-sense knowledge and various scientific disciplines, such as chemistry or history, are subject to major changes, however,

> ... no enthusiast of such a science or discipline is mad enough, or imperialist enough, to believe that it gives a complete description of its subject-matter. There are always other conjuncts to be had, typically from other disciplines.
>
> Not so, we submit, for physics, or at least theoretical physics: whether it by madness, imperialism, or part of what we mean by 'physics', physics *does* aspire to give just this sort of complete description of its subject-matter. And this implies that when 'other

[18] J. Butterfield and C.J. Isham, 'Spacetime and the Philosophical Challenge of Quantum Gravity', *Physics Meets Philosophy at the Planck Scale*, C. Callender, N. Huggett (eds.), (Cambridge, Cambridge University Press, 2001).

> conjuncts arrive' – i.e., new facts come to light additional to those given by the most complete available physical description – it is more reasonable to construe the new facts as involving a change of subject-matter, rather than as an additional piece of doctrine about the old subject-matter.[19]

To put this short: the subject-matter of physics presupposes a certain ontology of the world. If the subject-matter changes, the assumed ontology changes as well. And this implies that we should 'beware scientific realism'.[20]

The recent discussion in philosophy of physics, exactly on this problem, was initiated by John Worrall's paper entitled 'Structural Realism: The Best of Both Worlds?'.[21] He starts his analysis with the well-known argument on behalf of scientific realism, which he calls 'no miracles' argument:

> Very roughly, this argument goes as follows. It would be a miracle, a coincidence on a near cosmic scale, if a theory made as many correct empirical predictions as, say, the general theory of relativity or the photon theory of light *without* what that theory says about the fundamental structure of the universe being correct or 'essentially' or 'basically' correct.[22]

We thus have two conflicting views: realism and anti-realism, and both of them look for support in the progress of physics: accumulation of its results supports realism, and discontinuities in its description of the world supports anti-realism. Worrall asks:

> Is it possible to have the best of both worlds, to account (no matter how tentatively) for the empirical success of theoretical science without running foul of the historical facts about theory-change?[23]

We should distinguish the continuity or discontinuity in the evolution of physics at the level of empirical results and at the level of world's description. There are no major problems, as far as the 'essential cumulativity' at the empirical level is concerned; it is a 'non-cumulative kind at the top theoretical levels' that creates serious problems. One of the answers to these problems could be an instrumentalist view of science, according to which theories make 'no real claims beyond their directly empirical conse-

[19] *Ibid.*
[20] This is the title of a subsection in the above quoted paper by Butterfield and Isham.
[21] *Dialectica*, 43 (1989), pp. 97-124.
[22] *Ibid.*, p. 101.
[23] *Ibid.*, p. 111.

quences', but this is exactly what Worrall wants to avoid. The solution he proposes is called (by him) *structural realism*. He writes:

> The role in the history of physics seems to be that, whenever a theory replaces a predecessor, which has however itself enjoyed genuine predictive success, the 'correspondence principle' applies. This requires the *mathematical equations* of the old theory to re-emerge as limiting cases of the mathematical equations of the new.[24]

Even if there is a discontinuity in the 'world images' given by the old and new theories, there is a substantial continuity in their respective mathematical structures. It 'is not evidence for full-blown realism – but instead only for structural realism'.[25]

The change from Fresnel's wave theory to Maxwell's electromagnetic theory is an often discussed example in this context.[26] There is certainly a drastic 'ontological change' between light as a periodic disturbance in an elastic medium (Fresnel) and light as an excitation of the 'disembodied' field (Maxwell). However, the continuity is recovered as soon as we look at this 'theory shift' from the structuralist point of view.

> Fresnel's equations are taken over completely intact into the superseding theory – reappearing there newly interpreted but, as mathematical equations entirely unchanged.[27]

The latter situation is rather exceptional in the history of physics. A more common one, and entirely sufficient for the structuralist approach, is that the equations of the old theory appear as limiting cases of the new theory. Since mathematical equations describe structures, the example of Fresnel and Maxwell theories exhibits

> cumulative growth at the structural level combined with the radical replacement of the previous ontological ideas. It speaks, then, in favour of a *structural* realism.[28]

[24] *Ibid.*, p. 120.

[25] *Ibid.*, p. 121.

[26] It was already Poincaré who quoted this example (*Science and Hypothesis*, Dover, 1905, pp. 160-162) 'to argue for a general sort of *syntactic* or *structural realism* quite different from the anti-realist instrumentalism which is often attributed to him' (Worrall, op. cit., p. 117).

[27] Worrall, *op. cit.*, p. 120.

[28] *Ibid.*

4. What Is Structural Realism?

In his response to Warrall's paper, James Ladyman[29] quotes the structuralist views of Bertrand Russell and Grover Maxwell who claimed that

> the objective world is composed of unobservable objects between which certain properties and relations obtain; but we can only *know* the properties and relations of these properties and relations, that is the *structure* of the objective world.[30]

This view is qualified as an *epistemological* (or *epistemic*) *structuralism*. Ladyman himself opts for something more. He argues that

> structural realism gains no advantage over traditional scientific realism if it is understood as merely an epistemological refinement of it, and that instead it ought to be developed as a metaphysical position.[31]

Ladyman thinks that to cope with this problem one must change from the syntactic, or 'received', view to the semantic view on scientific theories. According to the syntactic view, scientific theory is but a system of propositions, in the ideal case, an axiomatic system suitably interpreted, whereas according to semantic approach

> theories are to be thought of as presenting structures or models that may be used to represent systems rather than partially-interpreted axiomatic systems. Theories are not collections of propositions or statements, but are 'extra-linguistic entities which may be described or characterised by a number of different linguistic formulations'.[32]

The distinction between epistemic structuralism and ontic structuralism was made sharp by Michael Esfeld who writes:

> If physics tells us only about the way in which things at the basic level of the world are related to each other, two different metaphysical positions are open: (1) The things at the basic level have intrinsic properties of which we cannot gain any knowledge insofar as

[29] 'What Is Structural Realism?', *Studies in the History and Philosophy of Science* 29, 1998, pp. 409-424.

[30] *Ibid.*, p. 412.

[31] *Ibid.*, p. 411.

[32] *Ibid.*, p. 416. The last sentence is a quotation from: F. Suppe, *The Structure of Scientific Theories* (Chicago, University of Illinois Press, 1974), p. 221.

> they are intrinsic. (2) The relations in which they stand are all there is to the things at the basic level.[33]

View (1) is called the *epistemic* (or *epistemological*) structuralism; view (2) is called the *ontic* (or *ontological*) structuralism.

In this paper, I am not so much concerned about the realism – anti-realism dispute, but rather about structuralist interpretation of physics. And from this point of view, the following conclusion should be regarded as well-founded: From the analysis of the mathematical-empirical method of physics it follows that physical theories are about structures of the world. Two interpretations of this fact are possible: either structures, discovered by physical theories, are structures of something, i.e., there is a structured stuff, but this stuff is transparent for the method of physics (epistemic version), or such a stuff is absent (ontic version). When adopting the second version, we could speak about objects between which structural relations obtain, but such objects should be understood as 'empty places' in a structure (empty – because devoid of any 'intrinsic properties'). To decide between these two versions, we should go beyond the analysis of the physical method, and appeal to metaphysical or logical reasons (for instance, to the claim that there cannot be structural relations without the relata), or to invoke some 'strategic principles' (like Ockham's rasor to claim that 'undetectable stuff' is not necessary).

5. Representation Invariants

An unavoidable question concerning structuralism is what should we understand by structure. An easy answer would be that the structure presupposed by a given physical theory should be identified with (or somehow correlated to) the mathematical structure this theory employs. Unfortunately, this answer meets the following difficulty. In physics, there are theories that admit more than one, apparently different, mathematical formulations. For instance, classical mechanics can be expressed in terms of the 'action at a distance', in terms of variational principles, in terms of space-time curvature or with the help of field theoretical methods. Quantum mechanics also admits several representations: in terms of operators on a Hilbert space, in

[33] M. Esfeld, 'Do Relations Require Underlying Intrinsic Properties? A Physical Argument for a Metaphysics of Relations', *Metaphysica*, 4 (2003), pp. 5-26.

terms of C^*-algebras, in terms of Feynman path-integrals, or in terms of density matrices. This creates the real problem for the traditional realism, but can be a chance for the structuralist interpretation; namely, we can claim that, in a case of many mathematical formulations of the same physical theory, we have various representations of *the same structure*. This was the view of Dirac and Weyl with regard to two different 'pictures' of quantum mechanics – the wave picture created by Schrödinger and the matrix picture created by Heisenberg. Ladyman seems to catch the correct solution:

> The idea then is that we have various representations which may be transformed or translated into one another, and then we have an invariant state under such transformations which represents the objective state of affairs.[34]

The idea goes back to the famous 'Erlangen Program' in geometry, formulated by Felix Klein in 1872. He classified all geometries existing at the time according to various symmetry groups, and defined a geometry as a theory of invariants with respect to a given group. This approach was taken up by Weyl in his analysis of relativity theory and quantum mechanics.

> For Weyl, appearances are open only for intuition (in the Kantian sense of subjective perception) and therefore agreement is obtained by giving objective status only to those relations that are invariant under particular transformations.[35]

The situation could be summarized in the following way. There is an abstract structure[36] which admits various representations: A, B, C, say. They are well defined mathematical structures. Let us suppose that these mathematical structures have been used to formulate the same physical theory, and all these formulations lead to correct empirical predictions. In such a case, all of them must have 'something in common', and consequently there must exist some 'translation rules' from A to B, from B to C, etc. If we change from one such representations to another, something must be preserved. Otherwise these mathematical structures could not have been representations of the same theory. It is not important whether we are able to identify these 'representation invariants', or not. The vital

[34] J. Ladyman, *art. cit.*, p. 421.

[35] *Ibid.*, p. 420.

[36] Which can be interpreted in a Platonic or anti-Platonic way; this is not the subject-matter of the present paper.

point is that they do exist. And my claim is (like that of Ladyman) that precisely the collection of these ‘representation invariants’ is what a given theory is about, what constitutes the *ontology* of this physical theory. The situation is not unlike when the meaning of a book is defined as an invariant with respect to the family of all its possible translations: indeed, the meaning of a book is what does not change when we change from one its (good) translation to another its (good) translation.

If we believe in the success of physics, we are entitled to claim that the structures of the successive physical theories (in the above sense) approximate, or at least are somehow related, to the Structure of the World.

PATH OF DISCOVERY OF A THERAPEUTIC VACCINE

MICHAEL SELA

Before discussing my specific case, I would like to approach the subject of 'paths of discovery' in a more general manner. Karl Popper in his many writings was of the opinion that the best way to contribute to building science is to progress by making conjectures and trying to refute them. A self-consistent model, even when it is wrong, is always useful if it is 'falsifiable' [1]. Disproving an accepted theory is the only way to advance knowledge. As Baruch Blumberg mentions in his discussion of the scientific process: 'not only should scientific experiments be planned in their technical aspects prior to starting, but their design should be part of a long range strategic plan' – all this because of the influence of Karl Popper's recommendations [2].

In my opinion, theories are very good as working hypotheses and as inducement to do some experiments, but they become terribly dangerous when they become dogmatic, and then they are also counterproductive. I mean theories are good as long as you do not take them too seriously. I must qualify this statement by one thing, because in the lab I want to do those experiments that excite and intrigue me: I do not want to have to spend a large part of the time just to disprove other people's theories or hypotheses. Sometimes you have a difficult situation because you cannot just say to somebody: 'No, I do not believe it, I am sure it is wrong'. Somebody has to test it and prove it, otherwise the scientific method falls apart.

I actually believe that one needs three things for research: optimism, perseverance, and serendipity, which is when luck meets the prepared mind. I also believe that some of the best ideas come totally unexpectedly during a time of relaxation, which could be either some wonderful bona fide relaxation, or some boring committee meetings.

I think it is very important to have good schooling, but it is desirable to have also a little bit of ignorance. I am a compulsive reader, or more precisely, a compulsive scanner of literature. But I think that if I had first decided to study immunology in depth for a couple of years as it was practiced at that time, and then had started asking questions, I do not know whether I would even have dared to start conducting experiments. My approach was that if I came to a complex biological reality, I should try to figure out whether I could define precisely one question which could be answered in a clear-cut way, and then we could continue and move to another question.

I would like to make another remark: I know there are individuals who really prefer to be alone all the time, and maybe through daydreaming, they reach all their working hypotheses, but I am a great believer in the interaction and fertilization of ideas. I feel there is a tremendous need and great payoff for this kind of interaction, and I have been collaborating around the world with many colleagues, and was always keen on having many visiting scientists from various parts of the world spending extended periods of time in my laboratory.

As an example from my own research, I would like to describe here the path of discovery of the drug, which I prefer to call the therapeutic vaccine, for the exacerbating-remitting form of multiple sclerosis [3, 4]. We called it 'copolymer 1'. It was renamed 'glatiramer acetate' and industry named it Copaxone. It all started with our interest in the immunology of lipids.

Copolymer 1 and Multiple Sclerosis

In our early studies with Ruth Arnon, of special interest was the immune response to lipid components, which was not easy to either elicit or investigate because of solubility problems. However, conjugates, in which synthetic lipid compounds were attached onto synthetic copolymers of amino acids, elicited a specific response to lipids such as cytolipin H, which is a tumor-associated glycolipid [5], or sphingomyelin [6]. Furthermore, we demonstrated that both the sugar and lipid components of such molecules contributed to their immunological specificity. The resultant anti-lipid antibodies were capable of detecting the corresponding lipids both in water-soluble systems and in their physiological milieu. This was fascinating because it gave us a glimpse into some disorders involving lipid-containing tissue and consequently led to our interest in demyelinating diseases, namely, disorders in which the myelin sheath, which constitutes the lipid-rich coating of all axons, is damaged, resulting in various neurological dysfunctions. We thus thought that EAE (experimental allergic encephalomyelitis) caused by

MBP (myelin basic protein) might actually be induced by a demyelinating lipid and that the positively charged MBP might serve only as a schlepper (carrier) for an acidic lipid (e.g. phospholipids). We prepared several positively charged copolymers of amino acids and tested to see whether we could induce EAE when the copolymers were administered into experimental animals (guinea pigs and rabbits) in complete Freund's adjuvant, similarly to the successful administration of MBP, but we failed. On the other hand, the injection of several positively charged amino acid copolymers in aqueous solution into mice, rabbits, and guinea pigs resulted in efficient suppression of the onset of EAE [7, 8]. Later we were able to suppress the actual disease in rhesus monkeys and baboons [9. 10]. The copolymer 1 we primarily used, denoted Cop 1, is composed of a small amount of glutamic acid, a much larger amount of lysine, some tyrosine, and a major share of alanine. To our pleasant surprise, there is a significant immunological cross-reaction (both at the antibody level [11, 12] and at the T cell level [13, 14], between Cop 1 and the myelin basic protein. Interestingly, when an analog of Cop 1 made from D-amino acids was tested, it had no suppressing capacity, nor did it cross-react immunologically with the basic protein [15]. Cop 1 is not generally immunosuppressive, nor is it toxic; actually it is not helpful in any other autoimmune disease except in multiple sclerosis and its animal model, experimental allergic encephalomyelitis.

The clinical trials with Cop 1 have included two preliminary open trials and two double-blind phase II trials, one involving exacerbating-remitting patients [16] and another one on chronic progressive patients [17]. The results of the phase II trial in exacerbating-remitting patients demonstrated a remarkable decrease in the number of relapses and in the rate of progression in Cop 1-treated patients compared with the placebo control. Cop 1 is a promising, low risk multiple sclerosis-specific drug for treatment of the relapsing disease. As an antigen-specific intervention, Cop 1 has the advantage of reduced probability of long term damage to the immune system.

After a successful, pivotal multicenter phase III clinical trial conducted in ll medical centers in the United States [18], Cop 1 was approved by the United States Food and Drug Administration as a drug for multiple sclerosis. This was a moment of gratification and deep emotion for my colleagues and myself, as well as for our industrial partners, Teva Pharmaceutical Industries.

We were obviously very interested in the mode of action of Cop 1. We knew that the effect was specific for the disease, and we assumed that it had to do with the immunological cross-reaction between the 'villain' (myelin basic protein) and the drug (Cop 1). What we learned later is that Cop 1 binds almost immediately and strongly to the groove of major histocompatibility complex (MHC) class II antigens of most genetic backgrounds, and it displaces efficiently from the groove any peptides derived from the myelin protein [19]. This promiscuity is probably because of its polymeric character, permitting microheterogeneity in the amino acid sequence. The extensive and promiscuous binding to class II MHC molecules, without prior processing, leads to clustering of these molecules on the antigen-presenting cells, which may explain their high affinity binding [20].

This is the first necessary but not sufficient step in its mechanism of action. The binding, which is the least specific step, is a prerequisite for its later effects. Following this interaction, two mechanisms were clearly shown to be effective. 1) Cop 1 binding to the relevant MHC leads to the activation of T suppressor cells because of suppressive determinants shared between myelin basic protein and Cop 1. 2) Successful competition between the complex of Cop 1-MHC class II antigen with the complex of myelin basic protein-MHC class II antigen for the myelin basic protein-specific T cell receptor (a phenomenon called by immunologists the 'T receptor antagonism') is shown [21].

An important step in our understanding of the mode of action of Cop 1 was the observation that copolymer 1 induces T cells of the T helper type 2 that cross-react with myelin basic protein and suppress experimental autoimmune encephalomyelitis [22]. This was corroborated by clinical studies in multiple sclerosis patients [23]. It was of interest to observe that Th2 suppressor lines and clones induced by Copolymer 1 cross-reacted at the level of Th2 cytokine secretion with myelin basic protein but not with other myelin antigens [24]. This bystander suppression may explain the therapeutic effect of Cop 1 in EAE and multiple sclerosis (MS).

Cop 1 binds promiscuously to many different cells regardless of their DR restriction. It binds avidly and fast and can also displace already bound antigens, and this holds for all the myelin antigens that may be involved in MS; and yet, Cop 1 exerts its activity in an antigen-specific manner (it is not a general immunosuppressive agent and does not affect other experimental autoimmune diseases). Its specificity must, therefore, be envisaged in the context of the trimolecular complex MHC-Ag-T-cell receptor ('the immunological synapse'), namely, as interference with the presentation of the encephalitogenic antigen to the T-cell receptor, which is a specific interaction.

I summarized recently the story of specific vaccines against autoimmune diseases [25], as well as the successful use of Cop 1 (glatiramer acetate, Copaxone) in the treatment of multiple sclerosis for exacerbating-remitting patients [26]. The majority of the patients in the large clinical trial have been followed in an organized fashion for more than 7 years. Their risk of an MS relapse was over 1.5 per year at onset and is now less than 1 every 6 years. On an average, these patients have experienced no increase in neurological disability, whereas natural history profiles would have predicted substantial worsening. The accumulated experience with glatiramer acetate (Cop 1) indicates that its efficiency is apparently increased as a function of usage time, while the favorable side effect profile is sustained.

Personally, the whole odyssey of Cop 1 and its use in MS has been a source of great satisfaction and emotion. The awareness that tens of thousands of MS patients feel better because of a drug/vaccine that we have conceived and developed, moves me deeply. Twenty-eight years have passed from the moment of the idea to the approval of Cop 1 by the Food and Drug Administration. I have a feeling that discoveries resulting from basic research take a longer time to fruition, but on the other hand, they are probably more original in terms of concept.

Therapeutic Vaccines

Copolymer 1 is just one example of a therapeutic vaccine in the field of autoimmunity. Great effort is being devoted to develop vaccines against tumors, AIDS, hepatitis, tuberculosis, and possibly against the bacteria that cause gastric ulcers. Copolymer 1, used today as a vaccine against multiple sclerosis (MS), is a good example of a beneficial treatment for this autoimmune disease, based on its similarity to the myelin basic protein (MBP), one of the putative causes of MS. This finding could lead to therapeutic vaccines against other autoimmune diseases such as myasthenia gravis, systemic lupus erythematosus (SLE) and rheumatoid arthritis. Furthermore, antibodies prepared against prions raise hopes for a vaccine against bovine spongiform encephalitis (BSE) and Creutzfeldt-Jakob disease (CJD) and antibodies to a peptide derived from b-Amyloid plaques could degrade plaques and be used as a therapeutic vaccine against Alzheimer's disease (AD) [27].

By its definition, a preventive vaccine is sufficiently similar in its chemistry to the molecule that provokes the disease so that the immune response directed against it can act against the causative agent. This situation is analogous in the case of therapeutic vaccines.

A colloquium on 'Therapeutic Vaccines' took place recently in the USA National Academy of Sciences in Washington, trying to put under one roof the manifold efforts in various areas in need of such a vaccine [28].

Conclusions

In conclusion, I feel that there must be some strategic planning of research; in my case this would be the molecular basis of antigenicity, synthetic vaccines, autoimmune diseases or cancer vaccines, but at the tactical level there must be space for spontaneity and serendipity, which I have already defined as a situation where luck meets the prepared mind.

REFERENCES

1. Dauchin, A., in A. Ullmann (ed.), *Origins of Molecular Biology*: *A Tribute to Jacques Monod* (Washington, ASM Press, 2003), p. 289.
2. Blumberg, B.S., *Creativity Research Journal*, 7, 315 (1994).
3. Sela, M., *J. Biol. Chem.*, 278, 48507 (2003).
4. Arnon, R., and Sela, M., *J. Mol. Recognition*, 16, 412 (2003).
5. Arnon, R., Sela, M., Rachaman, E.S., and Shapiro, D., *Eur. J. Biochem.*, 2 (1967).
6. Teitelbaum, D., Arnon, R., Rabinsohn, Y., and Shapiro, D., *Immunochemistry*, 10, 735 (1973).
7. Teitelbaum, D., Meshorer, A., Hirshfeld, T., Arnon, R., and Sela, M., *Eur. J. Immunol.*, 1, 242 (1971).
8. Teitelbaum, D., Webb, C., Meshorer, A., Arnon, R., and Sela, M., *Eur. J. Immunol.*, 3, 273 (1973).
9. Sela, M., Arnon, R. and Teitelbaum, D., *Bull. Inst. Pasteur*, 88, 303 (1990).
10. Teitelbaum, D., Webb, C., Bree, M., Meshorer, A., Arnon, R., and Sela, M., *Clin. Immunol. Immunopathol.*, 3, 256 (1974).
11. Webb, C., Teitelbaum, D., Arnon, R., and Sela, M., *Eur. J. Immunol.*, 3, 279 (1973).
12. Teitelbaum, D., Aharoni, R., Sela, M., and Arnon, R., *Proc. Natl. Acad. Sci. USA*, 88, 9528 (1991).
13. Teitelbaum, D., Aharoni, R., Arnon, R., and Sela, M., *Proc. Natl. Acad. Sci. USA*, 85, 9724 (1988).
14. Teitelbaum, D., Milo, R., Arnon, R., and Sela, M., *Proc. Natl. Acad. Sci. USA*, 89, 137, (1992).

15. Webb, C., Teitelbaum, D., Herz, A., Arnon, R., and Sela, M., *Immunochemistry*, 13, 333 (1976).
16. Bornstein, M.B., Miller, A., Slagle, S., Weitzman, M., Crystal, H., Drexler, E., Keilson, M., Merriam, A., Wassertheil-Smoller, S., Spada, V., Weiss, W., Arnon, R., Jacobsohn, I., Teitelbaum, D., and Sela, M., *N. Engl. J. Med.*, 37, 408 (1987).
17. Bornstein, M.B., Miller, A., Slagle, S., Weitzman, M., Drexler, E., Keilson, M., Spada, V., Weiss, W., Appel, S., Rolak, L., Harati, Y., Brown, S., Arnon, R., Jacobsohn, I., Teitelbaum, D. and Sela, M., *Neurology*, 41, 533 (1991).
18. Johnson, K.P., Brooks, B.R., Cohen, J.A., Ford, C.C., Goldstein, J., Lisak, R.P., Myers, L.W., Panitch, H.S., Rose, J.W., Schiffer, R.B., Vollner, T., Weiner, L.P., Wolinsky, J.S., and the Copolymer 1 MS Study Group, *Neurology*, 45, 1268 (1995).
19. Fridkis-Hareli, M., Teitelbaum, D., Gurevich, E., Pecht, I., Bautbar, C., Kwon, O.J., Brenner, T., Arnon, R., and Sela, M., *Proc. Natl. Acad. Sci. USA*, 91, 4872 (1994).
20. Fridkis-Hareli, M., Teitelbaum, D., Pecht, I., Arnon, R., and Sela, M., *Int. Immunol.*, 7, 925 (1997).
21. Aharoni, R., Teitelbaum, D., Arnon, R. and Sela, M., *Proc. Natl. Acad. Sci. USA*, 96, 634 (1999).
22. Aharoni, R., Teitelbaum, D., Sela, M., and Arnon, R., *Proc. Natl. Acad. Sci. USA*, 94,10821 (1997).
23. Neuhaus, O., Farina, C., Yassouridis, A., Wiendl, H., Bergh, F.T., Dose, T., Wekerle, H., and Hohlfeld, R., *Proc. Natl. Acad. Sci. USA*, 97, 7452 (2000).
24. Aharoni, R., Teitelbaum, D., Sela, M. and Arnon, R., *J. Neuroimmunol.*, 91, 135 (1998).
25. Sela, M., *C.R. Acad. Sci. Paris Life Sci.*, 322, 933 (1999).
26. Sela, M., and Teitelbaum, D., *Expert Opin. Pharmacother.*, 2, 1149 (2001).
27. Sela, M., Arnon, R., and Schechter, B., *Drug Discovery*, 7, 664 (2002).
28. Supplement to PNAS – Sackler 'Colloquium on Therapeutic Vaccines – Realities of Today and Hopes for Tomorrow', M. Sela and M. Hilleman (eds.), (2004).

THE HISTORY OF ASPIRIN: THE DISCOVERIES THAT CHANGED CONTEMPORARY MEDICINE

ANDREW SZCZEKLIK

Ancestors of Aspirin

For millennia, pain, fever, and inflammation were treated with plants containing salicylic acid glycosides: leaves of myrtle (Myrtus), bark of willow (Salix), bark of poplar (Populus), meadowsweet (Spirea). About 3500 years ago, an Egyptian papyrus recommended the application of a decoction of the dried leaves of Myrtle to the abdomen and back to expel rheumatic pains from the womb. A thousand years later, Hippocrates championed the juices of the poplar tree for treating eye diseases, and those of willow bark for pain in childbirth and for fever. Through the Middle Ages the medical use of salicylates continued. However, willows were needed for basket-making so the women grew meadowsweet (Spirea ulmaria) in their gardens and made decoctions from the flowers [1].

The first 'clinical trial' of willow bark was published by English country parson Reverend Edward Stone. He had accidentally tasted willow bark and was surprised by its bitterness, which reminded him of chinchona bark – containing quinine – then being used to treat malaria. He believed in the 'doctrine of signature' which dictated that the cures for the diseases would be found in the same location where malady occurs. Since the 'Willow delights in a moist and wet soil, where agues chiefly abound', he gathered a pound of willow bark, dried it in a baker's oven for 3 months, pulverized and administered it to 50 patients with safety and success. He reported his observations to the Royal Society [2].

In 1859 Friedrich Kolbe identified the structure of salicylic acid as an o-hydroxybenzoic acid, managed to obtain it synthetically and introduced it to therapy. However, the bitter taste of the substance and the side-effects of gastric irritation caused by acid, made long-term prescribing difficult. In

1897 Felix Hoffman at Bayer's Laboratory synthesized acetylsalicylic acid, shortly thereafter named 'aspirin' (it contained the root of spiric acid from Spirea Ulmaria, chemically identical to salicylic acid, together with 'A' as an abbreviation for acid) (Fig. 1). Hoffman had personal reasons for wanting a more acceptable salicylic acid derivative, since his father who had been taking salicylic acid for many years to treat his painful arthritis had recently discovered that he could no longer take the drug without vomiting [1].

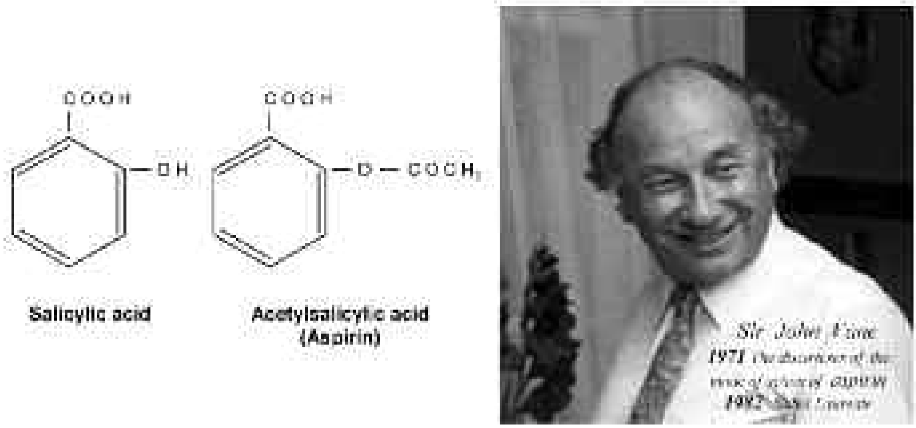


Fig. 1 Chemical structures of salicylic – and acetyl-salicylic acids.

Use of Aspirin and its Mode of Action

Following its introduction into therapy at the beginning of 20th century, aspirin has become the extremely popular anti-pyretic, anti-inflammatory and analgesic agent. The crucial discovery of the biochemical mechanism of the action of aspirin was made by Sir John Vane in 1971 [3]. He observed that aspirin blocked the enzymatic activity of cyclooxygenase (COX), a key enzyme leading to the production of pro-inflammatory prostaglandins from arachidonic acid (Fig. 2). He thus demonstrated the reasons for the anti-inflammatory, analgesic, antipyretic and toxic effects of the most widely used remedy of all time. The discovery in part

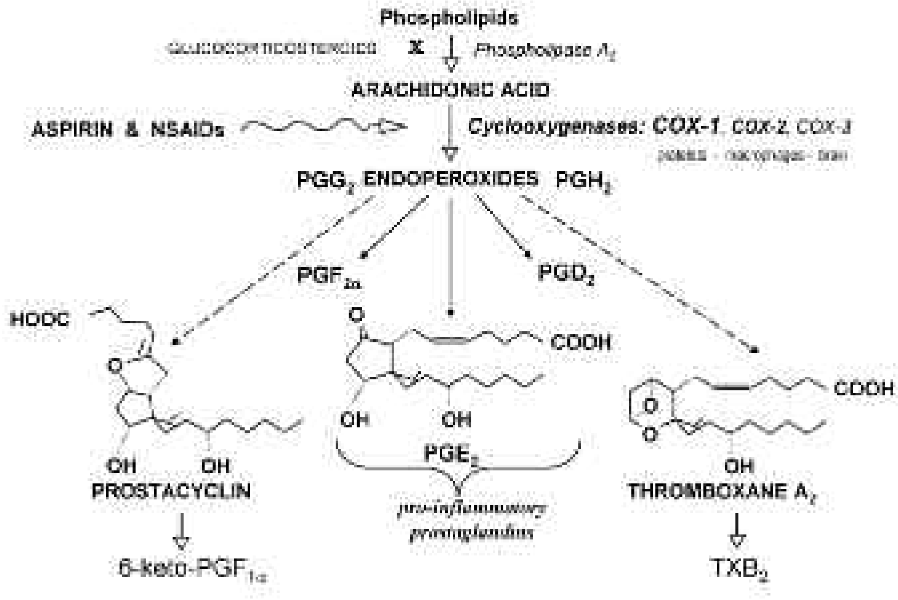


Fig. 2 Arachidonic acid metabolism, controlled by cyclooxygenases. Arachidonic acid is liberated from the cells by phospholipase A_2, and converted by cyclooxygenases via unstable endoperoxides to prostaglandins. In endothelial cells prostacyclin is preferentially produced, while blood platelets generate thromboxane A_2.

coincided with a veritable explosion of aspirin-like drugs, exemplified by phenylobutazone, indomethacin, mefenamic acid, ibuprofen and others, widely used for treating rheumatic condition. All of the so-called non-steroidal anti-inflammatory drugs (NSAIDs) inhibited prostaglandin biosynthesis, but their toxicity remained a problem and was probably caused by the same biological intervention.

Over the last twenty years, yet another use for aspirin has emerged connected with the discovery of its anti-thrombotic action [4]. The peculiar effectiveness of aspirin in thrombosis is based on its irreversible inhibition of the COX enzyme in platelets which make thromboxane A_2 (TXA_2), the potent pro-aggregatory and vasoconstrictive prostaglandin. Of all the aspirin-like drugs, only aspirin permanently inactivates the enzyme (by attaching an acetyl group close to the active site), and thus makes the platelet incapable of synthesizing TXA_2 for the rest of its life (about 10 days). Regular low doses of aspirin have a cumulative effect and inactivate platelet COX without affecting much the synthesis in blood vessels of prostacyclin – which exert potent vas-dilatory and anti-platelet effects [5].

Effectiveness of Aspirin in Cardiovascular Disease

Aspirin is the cornerstone of therapy in atherothrombosis, encompassing a wide spectrum of clinical entities. A recent meta-analysis of 287 clinical trials on aspirin in the prevention of cardiovascular disease has provided firm evidence that antiplatelet therapy, mainly aspirin, can reduce by approximately 25% the risk of nonfatal myocardial infarction (MI), nonfatal stroke, or vascular death in high-risk patients, regardless of sex, age, presence of arterial hypertension or diabetes [6]. An absolute reduction in the risk of having a serious vascular event was 36 per 1000 MI survivors treated for two years [6]. Undoubtedly, the clinical benefits of aspirin are most apparent in patients with acute myocardial infarction, which has been convincingly demonstrated in the landmark infarction trial, ISIS-2 [7]. Administration of 162mg aspirin within 24 hours following the onset of acute coronary symptoms resulted in significant reductions in the risk of cardiovascular mortality (by 23%), nonfatal reinfarction (by 49%), and nonfatal stroke (by 46%) in the 5-week follow-up, without any increase in major bleeding complications [7]. The magnitude of the effect attributed to aspirin was almost equivalent to that produced by streptokinase [7]. This study showed that aspirin has the best benefit-to-risk ratio of any known therapy for AMI. Aspirin is also effective in reducing the risk of MI or sudden death by about 30% in patients with stable angina [8]. Current state of evidence favors the use of aspirin as the first-line agent for the majority of patients with vascular disease [9]. Although recommended by existing guidelines in diabetes, its efficacy appears to be substantially lower than in non-diabetic individuals [10]. Data on a role of aspirin among low-risk patients are inconsistent. In 2001, the first randomized controlled trial, PPP (the Primary Prevention Project) [11], conducted in men and women aged 50 years or more, provided the direct evidence of aspirin's efficacy in prevention of cardiovascular events such as angina, peripheral artery disease, and transient ischemic attacks (TIA) and demonstrated a reduction in the relative risk of death to 0.56 among individuals taking 100mg aspirin daily. Based on the meta-analysis of four primary prevention trials, including the

US Physicians' Health Study and the British Doctors' Trial, it has been concluded that the use of low-dose aspirin is safe and effective in subjects with coronary event risk of at least 1.5 % per year [12].

Both European and American guidelines recommend aspirin administration at a daily dose of 75 to 100mg in high-risk patients [13, 14]. In contrast, a small absolute benefit in terms of cardiovascular morbidity and mortality does not justify a routine use of aspirin among low- risk individuals, in whom potential protective effects could be offset by exposure of thousands of apparently healthy individuals to hemorrhagic complications, and in asthmatics to aggravation of their disease [14, 15]. The FDA has not approved aspirin for primary prevention of cardiovascular disease.

Aspirin Resistance

Treatment failures occur with any drug and aspirin is no exception. Evidence is growing to indicate there are subpopulations that do not respond to the antithrombotic action of aspirin. The term 'aspirin resistance' has been used to describe a number of different phenomena, including inability of aspirin to: 1) protect against cardiovascular events despite its regular intake; 2) affect various laboratory tests, reflecting platelet activity. Research on aspirin resistance yielded interesting results in clinical pharmacology and pharmacogenetics [11, 16]. Future studies will show whether genotyping for polymorphisms might be of value in everyday clinical use of aspirin. Present data indicate that in survivors of recent myocardial infarction or unstable angina, patients receiving coronary artery bypass grafts, as well as in subjects with hypercholesterolemia, aspirin resistance has to be considered when implementing anti-thrombotic therapy (table 1). However, in individual patients the available laboratory tests are of no particular use to predict reliably the clinical outcome or to guide in making therapeutic decision. Prospective clinical trials seem necessary to reach such conclusions.

Attempts to Develop a Better Aspirin

Although aspirin and other NSAIDs effectively relieve symptoms, such as pain, the relief comes at the expense of important side effects, most notably upper gastrointestinal toxicity and exacerbation of existing asthma [14, 15] (Fig. 3). So when in early 1990, it was discovered that there

TABLE 1. Possible mechanisms of aspirin resistance

Reduced bioavailability

- Poor compliance[20]
- Concurrent intake of certain NSAIDs[21]

Enhanced platelet function

- Hypercholesterolemia, usually accompanied by increased thrombin generation[22]
- Hypercoagulable states following MI[23] and unstable angina[24]
- Biosynthesis of TXA_2 by pathways that are not blocked by aspirin, e.g. by COX-2 in monocytes and macrophages[25]
- Increased release of platelets from bone marrow in response to stress, i.e. after coronary artery bypass surgery[25]
- Transcellular arachidonate metabolism between aspirinated platelets and endothelial cells[26]

Genetic polymorphisms

- Polymorphism of platelet glycoprotein IIb/IIIa; carriers of PlA^2 allele are less sensitive to anti-thrombotic action of aspirin in vivo[27,28]
- COX-2 variants in patients after coronary artery bypass surgery[29]

Other factors

- Smoking[30] or increased levels of noradrenaline (excessive physical exercise, mental stress)[31]

are two COXs, COX-1 and COX-2, coded by two different genes, an effort was made to design drugs that would inhibit specifically COX-2. Their development was based on the hypothesis that COX-2 was the source of prostaglandins E_2 which mediates inflammation, and that COX-1 was the source of the same prostaglandins in gastric epithelium, where they afford cytoprotection. Indeed, coxibs, as they came to be called, proved to be much safer on gastro-intestinal tract than aspirin and other NSAIDs; moreover, they were perfectly well tolerated by adult asthmatics in whom aspirin precipitated asthmatic attacks [17, 18]. Coxibs have been aggres-

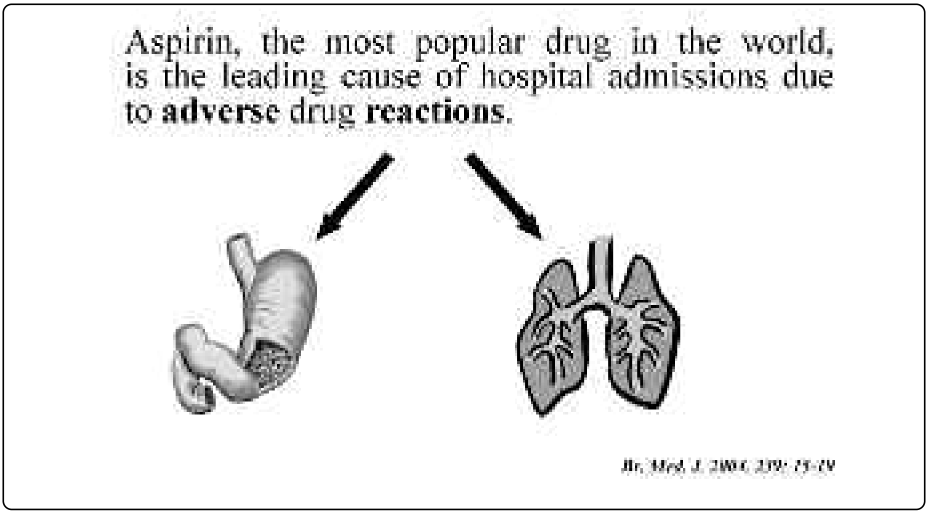


Fig. 3. Common side-effects of aspirin therapy: gastric ulcer and exacerbation of existing asthma.

sively marketed and have rapidly dominated the prescription-drug market for NSAIDs, accounting for worldwide sales of roughly $10 billion. However, in 2004, five years after its introduction, the leading coxib, rofecoxib (VIOXX), was withdrawn because its use was associated with unexpected but significant increase in myocardial infarction. The most likely explanation of this serious side-effect is that rofecoxib and probably other coxibs suppress the formation of prostacyclin in the endothelial cell of arteries, without affecting TXA_2 production in platelets. Thus, a single mechanism, depression of prostacyclin might be expected to elevate blood pressure, accelerate atherogenesis and predispose patients receiving coxibs to an exaggerated thrombotic response and to the rupture of an atherosclerotic plaque. How could it happen? Why did it take five years of the drug on the market to realize its dangerous side-effects? When the drug was introduced it was assumed that prostacyclin was derived mainly from COX-1 in endothelial cells. This assumption later proved incorrect, since studies in mice and humans showed that COX-2 was the dominant source [19]. The rofecoxib story also reflects poorly on the process that leads to drug approval [17].

SUMMARY

Slightly one hundred years after its introduction to therapy, aspirin remains the most popular drug in the world. Its medical indication have been increasing constantly, and now cover large areas beyond inflammation, including anti-thrombotic effects, and most recently anti-cancer activity. Yet, the recent story with coxibs teach us how careful we have to be in trying to improve actions of an extraordinary and well-known drug such as aspirin.

REFERENCES

1. Vane, J.R., and Botting, R.M. (eds.), *Aspirin and other salicylates* (Chapman and Hall, London, 1992).
2. Stone, E., 'An account of the success of the bark of the willow in the cure of agues', *Phil. Trans. R. Soc.*, London, 53:195-200 (1763).
3. Vane, J.R., 'Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs', *Nat. New. Biol.*, 231:232-5 (1971).
4. Szczeklik, A., Musial, J., Undas, A., Sanak, M., Dropinski, J., Tuleja, E., Wegrzyn, W., 'Aspirin and thrombinogenesis', *Thromb. Res.*; 110:345-7 (2003).
5. Szczeklik, A., Niżankowski, R., Skawiński, S., Szczeklik, J., Głuszko, P., Gryglewski, R.J., 'Successful therapy of advanced arteriosclerosis obliterans with prostacyclin', *Lancet*, 1:1111-1114 (1979).
6. Antithrombotic Trialists Collaboration Collaborative meta-analysis of randomized trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients, *BMJ*, 324:71-86 (2002).
7. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group, Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2, *Lancet*, 2:349-360 (1988).
8. Juul-Moller S., Edvardsson N., Jahnmatz B., Rosen A., Sorensen S., Omblus R., 'Double-blind trial of aspirin in primary prevention of myocardial infarction in patients with stable chronic angina pectoris', *Lancet*, 340:1421-1425 (1992).
9. Tran H., Anand S.S., 'Oral antiplatelet therapy in cerebrovascular disease, coronary artery disease, and peripheral arterial disease', *JAMA*, 292:1867-1874 (2004).
10. Evangelista V., Totani L., Rotondo S., Lorenzet R., Tognoni G., De Berardis G., Nicolucci A., 'Prevention of cardiovascular disease in type-

2 diabetes: How to improve the clinical efficacy of aspirin', *Thromb. Haemost*, 93:8-16 (2005).

11. de Gaetano G., 'Collaborative Group of the Primary Prevention Project (PPP). Low-dose aspirin and vitamin E in people at cardiovascular risk: a randomized trial in general practice', *Lancet*, 357:89-95 (2001).
12. Sanmuganathan P.S., Ghahramani P., Jackson P.R., Wallis E.J., Ramsay L.E., 'Aspirin for primary prevention of coronary heart disease: safety and absolute benefit related to coronary risk derived from meta-analysis of randomized trials', *Heart*, 85:265-271 (2001).
13. Patrono C., Bachmann F., Baigent C., Bode C., De Caterina R., Charbonnier B., Fitzgerald D., Hirsh J., Husted S., Kvasnicka J., Montalescot G., Garcia Rodriguez L.A., Verheugt F., Vermylen J., Wallentin L., Grupo de Trabajo sobre el uso de agentes antiplaquetarios en pacientes con enfermedad cardiovascular aterosclerotica de la Sociedad Europea de Cardiologia. Expert consensus document on the use of antiplatelet agents. The Task Force on the use of antiplatelet agents in patients with atherosclerotic cardiovascular disease of the European Society of Cardiology, Eur. *Heart J.*, 25:166-181 (2004).
14. Smith S.C. Jr, Blair S.N., Bonow R.O., Brass L.M., Cerqueira M.D., Dracup K., Fuster V., Gotto A., Grundy S.M., Miller N.H., Jacobs A., Jones D., Krauss R.M., Mosca L., Ockene I., Pasternak R.C., Pearson T., Pfeffer M.A., Starke R.D., Taubert K.A., 'AHA/ACC guidelines for preventing heart attack and death in patients with atherosclerotic cardiovascular disease: 2001 update. A statement for healthcare professionals from the American Heart Association and the American College of Cardiology', *Circulation*, 104:1577-1579 (2001).
15. Szczeklik A., Stevenson D.D., 'Aspirin-induced asthma: advances in pathogenesis, diagnosis, and management', *J. Allergy Clin. Immunol.*, 111:913-921 (2003).
16. Szczeklik A., Musiał J., Undas A., Sanak M., 'Aspirin resistance', *J. Thromb. Haemost.* (2005), in press.
17. Maxwell S.R., Webb D.J., 'COX-2 selective inhibitors – important lessons learned', *Lancet*, 365:449-51 (2005).
18. Szczeklik A., Niżankowska E., Bochenek G., Nagraba K., Mejza F., Świerczyńska M., 'Safety of a specific COX-2 inhibitor in aspirin-induced asthma', *Clin. Exp. Allergy*, 31:219-225 (2001).
19. Fitzgerald G.A., 'Coxibs and cardiovascular disease', *N. Engl. J. Med.*, 351:1709-11 (2004).
20. Hankey G.J., Eikelboom J.W., 'Aspirin resistance', *BMJ*, 328:477-479 (2004).

21. Catella-Lawson F., Reilly M.P., Kapoor S.C., Cucchiara A.J., DeMarco S., Tournier B., Vyas S.N., FitzGerald G.A., 'Cyclooxygenase inhibitors and the antiplatelet effects of aspirin', *N. Engl. J. Med.*, 345:1809-1817 (2001).
22. Szczeklik A., Musiał J., Undas A., Swadźba J., Góra P.F., Piwowarska W., Duplaga M., 'Inhibition of thrombin generation by aspirin is blunted in hypercholesterolemia', *Arterioscl. Thromb. Vasc. Biol.*, 16:948-954 (1996).
23. Szczeklik A., Dropinski J., Radwan J., Krzanowski M., 'Persistent generation of thrombin after acute myocardial infarction', *Arterioscler. Thromb.*, 12:548-553 (1992).
24. Merlini P.A., Bauer K.A., Oltrona L., Ardissino D., Cattaneo M., Belli C., Mannucci P.M., Rosenberg R.D., 'Persistent activation of coagulation mechanism in unstable angina and myocardial infarction', *Circulation*, 90:61-68 (1994).
25. Rocca B., Seccheiero P., Ciabattoni G., Ranelletti F.O., Catani L., Guidotti L., Melloni E., Maggiano N., Zauli G., Patrono C., 'Cyclooxygenase-2 expression is induced during human megakaryopoiesis and characterizes newly formed platelets', *Proc. Natl. Acad. Sci. USA*, 99:7634-7639 (2002).
26. Karim S., Habib A., Levy-Toledano S., Maclouf J., 'Cyclooxygenase-1 and -2 of endothelial cells utilize exogenous or endogenous arachidonic acid for transcellular production of thromboxane', *J. Biol. Chem.*, 271:12042-12048 (1996).
27. Szczeklik A., Undas A., Sanak M., Frolow M., Wegrzyn W., 'Relationship between bleeding time, aspirin and the PlA1/A2 polymorphism of platelet glycoprotein IIIa', *Br. J. Haematol.*, 110:965-967 (2000).
28. Undas A., Brummel K., Musial J., Mann K.G., Szczeklik A., 'Pl(A2) polymorphism of beta (3) integrins is associated with enhanced thrombin generation and impaired antithrombotic action of aspirin at the site of microvascular injury', *Circulation*, 104:2666-2672 (2001).
29. Censarek P., Freidel K., Udelhoven M., Ku S-J., Hohlfeld T., Meyer-Kirchrath J., Schroer K., Weber A-A., 'Cyclooxygenase COX-2a, a novel COX-2 mRNA variant, in platelets from patients after coronary artery bypass grafting', *Thromb. Haemost.*, 92:925-928 (2004).
30. Sane D.C., McKee S.A., Malinin A.I., Serebruany V.L., 'Frequency of aspirin resistance in patients with congestive heart failure treated with antecedent aspirin', *Am. J. Cardiol.*, 90:893-895 (2002).
31. Mustonen P., van Willigen G., Lassila R., 'Epinephrine – via activation of p38-MAPK – abolishes the effect of aspirin on platelet deposition to collagen', *Thromb. Res.*, 104:439-449 (2001).

THE ROLE OF INNOVATION, INTERDISCIPLINARITY AND PHENOMENOLOGY AS COMPONENTS OF CREATIVITY IN OPENING NEW WINDOWS

M.G.K. MENON

Introduction

Discovery involves the opening of a window to gain knowledge concerning the hitherto unknown. But the unknown, now being perceived through the discovery, already exists. This is an attribute fundamental to the concept of discovery. There are many ways in which discoveries are made. The aim of this particular meeting is to examine the multitudes of paths that lead to discovery, as also to look at some case studies based on personal experience.

In this paper, I wish to bring out the role that innovation, interdisciplinarity and phenomenology play in the process of discovery. There are many attributes associated with those who make discoveries: curiosity, originality, creativity, dedication and persistence, intelligence. The actual discovery may not be the direct result of a particular effort that is made, but would, in some sense, relate to it. The important words associated with the processes of discovery are: 'why'; 'what'; 'how'; – rather than 'who', which currently dominates society and media attention. Equally, it must be pointed out that just as 'why' is important so is, 'why not'.

Importance of a Questioning Environment for Discovery

Many of the older and static societies are dominated by dogma, by hierarchy and blind acceptance of authority. In this situation it is difficult for any individual to ask, on a free basis, any questions that would be in conflict with the existing situation. No doubt, in such societies, it is possible to ask all questions that do not in any way impinge on the existing order.

Concerning this, Gautama, the Buddha, had remarked:

> Believe nothing
> merely because you have been told it
> or because it is traditional
> or because you yourself have imagined it
> Do not believe what your teacher tells you,
> Merely out of respect for the teacher
> But whatever after due examination and analysis
> You find to be conducive to the good, the benefit,
> The welfare of all beings,
> That doctrine believe and cling to,
> And take it as your guide.

Role of Creativity

Another important element of discovery is creativity. This is manifest in many areas: science, engineering, architecture, art (painting, sculpture), performing arts, philosophy, religion, and many other such activities. Different parts of the brain are involved in the manifestation of creativity in these different areas. Creativity involves aspects that are different from pure intelligence. The latter is based on the faculty of thought and reason. Thus, through intelligence one may acquire and apply knowledge, and make extensions beyond existing knowledge, through the power of analytical reasoning. In the case of creativity, however, one deals with knowledge which has in it aspects of intuition, and generates new knowledge for which the building blocks are not the existing knowledge. One may use the same tools involved in analytical reasoning, or in areas of professional competence, but the result is something significantly related to imagination. Einstein remarked: 'Imagination is more important than knowledge'. While certainly there are many who are highly intelligent and also highly creative, one has many other instances of those not so intelligent, but who are highly creative. Apart from intelligence, which is undoubtedly a valuable asset for creativity, there are the very important attributes of dedication and commitment that are important for discovery.

Importance of an Interactive Environment

Concerning the role of an overall supportive environment and of team work, Jacques Monod commented in his Nobel Lecture on the importance to him of a Rockefeller Fellowship which gave him an opportunity to work at the California Institute of Technology in the laboratory of the Nobel Laureate Morgan:

> This was a revelation to me – a revelation of what a group of scientists could be like when engaged in creative activity, and sharing it in constant exchange of ideas, bold speculations and strong criticisms.

It is very clear that most human beings never stretch themselves to the limits of their abilities. It is the outstanding teacher who attracts the finest students; and, in the overall intellectual environment that the team represents, the individuals are pushed to the very limits of their intellectual capability, each deriving strength from the other in a resonant and supportive manner.

Illumination

New ideas often come to those who have been constantly thinking and struggling with a problem. This is beautifully illustrated by the great mathematician, Henri Poincare, who said:

> most striking at first is this appearance of sudden illumination, a manifest sign of long unconscious prior work. The role of this unconscious work in mathematical innovation appears to be incontestable.

Poincare gave a striking example:

> One evening, contrary to my custom, I drank black coffee and could not sleep. Ideas rose in crowds. I felt them collide until pairs interlocked so to speak, making a stable combination.

This related to Poincare's greatest discoveries, the theory of fuchsian groups and fuchsian functions, and arithmetic transformations of indefinite ternary quadratic forms.

Again, another great mathematician, Gauss, remarked:

> finally two days ago I succeeded, not on account of my painful efforts, but by the grace of God. Like a sudden flash of lightening the riddle happened to be solved.

Closer to home, in India, there was the mathematical genius Srinivasa Ramanujan. About him, the Cambridge mathematician, G.H. Hardy, remarked:

> In his favorite topics like 'Infinite Series and Continued Fractions' he had no equal in this century. His insight into the algebraic formulae, often (and unusually) brought about by considering innumerable examples, was truly amazing.

At the age of 16, he came across a book by George Carr which introduced him to the real world of mathematics. In this book only results were given, and proofs relegated to footnotes. Ramanujan proved the theorems by himself, often just as he got up in the morning. He claimed that the Goddess of Namakkal, (the goddess of learning) the family deity, had inspired him with the proofs in his dreams. He went through his short life, stating innumerable highly original and unconventional mathematical propositions in number theory, modular function theory and infinite theory on which mathematicians are still working; he did not provide any proofs. He went on to insist that none of these were his creation but revealed to him in dreams by the goddess.

According to many psychologists, creative thinking involves multiple stages: of preparation relating to the problem of interest and related aspects; incubation, whilst the mind is allowed to work on the problem in the subconscious; illumination which has been exemplified in the above examples of Poincare, Gauss and Ramanujan; and finally what is characteristic of the scientific method, the processes of analysis, validation and verification.

Invention, Creativity and Innovation

It has been pointed out already that discovery involves the process of perceiving that which already exists, whether it be a phenomenon or a law of nature. Children are discovering the world around them all the time; this is through their innate sense of curiosity, and the need for that knowledge for the daily processes of life and living. How and when that sense of curiosity dies down in most and changes take place in the cognitive processes needs further study. Information technology can provide a creative interactive environment, but an information overload may be quite negative.

However, when we talk of scientific discoveries, we imply that which has not been discovered before; and, therefore, truly constitutes a new window on nature.

In contrast, an invention involves the creation of something new – an artefact which has not existed before. This is a fundamental difference between discovery and invention. Because of the importance now attached to knowledge as a property right, inventions are very well defined in a legal sense, and protected by intellectual property rights.

Innovation is also defined as ‘introducing something new’. It involves a process of change and of novelty; hopefully it should lead to successful exploitation of the new ideas. Innovation is often empirical, a result of tinkering, namely trying to do something in different ways, and from these efforts the best way would emerge. There is clearly intuition, creativity and intelligence, apart from originality, involved in innovation.

As different from scientific discoveries and inventions (that are largely in the area of artefacts and other aspects covered by intellectual property and copyrights law), innovation has a much broader connotation. Innovation may lead to a discovery or result in an invention. But innovation is much broader, extending over all aspects of human activity and at all levels, from the daily small opportunities to unique ones. Thus one talks today of innovation that relate to sectors such as management, marketing, human resources, law and much else. Very often, innovations that finally result in the application of an idea on a large scale, with an impact on society, has to take place across the full innovation chain, involving all or most of the facets just mentioned.

Internet was the result of an effort to achieve a specific objective: to network computer systems to achieve resilience in the case of a nuclear attack. This effort was initiated in the USA by DARPA (Advanced Research Projects Agency of the Department of Defence, USA). It involved MIT, Stanford University, Carnegie-Mellon University and others. When accomplished, it was taken over and used by a limited number of academic institutions. We know the pioneers who worked on it. It was not a sudden discovery; it was something that evolved. There is a humorous remark that it was the finest innovation ever made by a Committee! When the National Science Foundation took it over, and the Internet Protocol was developed, it became universal. And when the World Wide Web was developed at CERN by Tim Berners-Lee, Internet became what it is: an ever expanding horizon of global uses and users. The final outcome is the backbone of the IT revolution: the pioneering discovery of the transistor, and later the invention of the integrated circuit, have resulted in Nobel Prizes to those concerned. Otherwise, only components of this edifice, particularly areas of application, have been rewarded with intellectual property rights. But, overall, this was a grand innovation in the area of public good.

In 1992, one of the great industrial leaders of Japan, Akio Morita, Chairman, Sony Corporation, delivered the first United Kingdom Innovation Lecture at The Royal Society in London. The title of his lecture was: '"S" does not equal "T" and "T" does not equal "I"'. What he had to say was that scientific research provides us with information concerning that which was previously unknown; it provides new windows to look out at natural phenomena; and can result in great discoveries. But this, by itself, though it involves a high degree of creativity and originality, does not necessarily lead to applications or the development of technology. Technology is a process involving the manipulation of science and its discoveries to lead to concepts; processes and devices. He stated that technologists have a role to play not only in the development of technology, but also in leading high technology and manufacturing businesses. But he went on to say that technology alone is not innovation. Whilst there was need for creativity in evolving new technology e.g. involving new concepts, processes and devices, it was also necessary to have associated creativity in product planning and creativity in marketing, to enable the new technology to fructify. He further emphasized the importance of innovation in management, and the importance of an innovation mandate which has to be defined by business as well as by government. Morita was referring to innovation in the context of the totality of the 'innovation chain' which has to be effective for successful or commercialization and large scale application to take place.

In a personal discussion he argued why the pocket (transistor) radio was not a great discovery or technological invention, but was an important innovation. The valve radio was known and in use. The transistor had been discovered; thereafter the integrated circuit had been invented. When these were used to replace the valve, the circuits became smaller and lighter, with less demand on power. Small dry cells could be used. The loudspeaker had to be miniaturized and the whole entity packaged and marketed. There were a large number of innovations which added up to make a huge commercial success.

INTERDISCIPLINARITY

In the early days of scientific development, science was regarded as natural philosophy – and that designation still persists in some educational institutions. That was when science concerned itself purely with theoretical and philosophical questions. This changed when the experimental method became part of science. Even at that stage, with the limited number of extraordinary individuals engaged in the scientific endeavour, whose names have come down to us – the rest being purely technical practitioners – their coverage was wide, Thus a Galileo or Newton covered areas of mechanics, optics, gravitation, astronomy and mathematics. A Faraday would deal with electricity, magnetism, chemistry and the like. Where can there be another Leonardo da Vinci who spanned science, engineering and creative arts at his level? More individuals could be cited; and even at other levels, they were truly natural philosophers. With the passage of time, this has changed. Many great names were associated with narrower fields in which major advances took place. Then, as understanding increased, fields started to come together. One of the characteristics of modern science has been the process of bringing fields together, through a broader understanding of common principles. Thus, all forms of energy began to be understood in a unified frame work: the law of conservation of energy and thermodynamics were born. Electricity and magnetism converged within the framework of electromagnetism (em). Electrodynamics developed. The em and weak forces were seen to be components of an electroweak interaction, and then chromodynamics which unified all of the strong interactions. From what is the 'standard model', one is looking for a 'theory of everything'. Under the rubric of the periodic table, all of the elements that constitute chemistry came together. With atomic theory and the chemical bond there arose an understanding of chemical reactions. Electrical energy was the first of the wholly science-based technologies. Now chemistry and chemical engineering have also changed from being empirical areas to being wholly science-based.

In general, there is a move towards convergence in technologies, (the most striking being in information technology and biotechnology), and of unification in science. This has, and will continue to offer opportunities for cross-over of ideas and techniques from one field to the other. Even more, such crossovers can take place, with revolutionary consequences, between highly disparate fields.

To take one area at the frontiers of modern science – that of the brain and neurosciences. This is a field where classical neurosciences (neurology, neurosurgery, psychiatry etc.) intersect now with molecular and cellular biology, biochemistry, cognition and sensory perception, education, psychology, imaging techniques, mathematics, computer sciences and artificial intelligence, engineering and so on. This is getting to be truly inter disciplinary – perhaps a fairly unique example.

Whilst discoveries will continue to be made in specific narrow disciplines, it is at the newly developing interfaces of the various disciplines that the greatest opportunities lie.

PHENOMENOLOGY

Nature is studied, first and foremost, by observations relating to the various ways in which it manifests itself; phenomena are those that are perceived by the senses and augmented by devices that enhance the power of the senses. The scientific study of phenomena is referred to as 'phenomenology'. Many of the great discoveries in science have come about through phenomenological studies.

Many in society regard these discoveries as accidental, or attribute them to serendipity namely, making discoveries of things that one is not in quest of, by accident and sagacity. There are many instances that one can give in this regard: for example, the discovery by Henri Becquerel of photographic material that had blackened sitting in a drawer – which opened up a whole new field of radioactivity, leading to nuclear physics and all that followed; or the discovery, by Alexander Fleming, of penicillin in an open glass dish, which opened up the whole field of antibiotics. These events are not accidents in the standard sense of that term. The discoverers were individuals with a sense of curiosity, sagacity, originality and, later, dedication and commitment to pursue the lead, so necessary for any such discovery to be meaningful.

I would like to recount what Karl Popper had said in this connection:

> Suppose that someone wished to give his whole life to Science. Suppose that he therefore sat down, pencil in hand, and for the next twenty, thirty, forty years recorded in notebook after notebook everything that he could observe. He may be supposed to leave out nothing: today's humidity, the racing results, the level of cosmic radiation, the stock market prices and the look of Mars, all would be there. He would have compiled the most careful record of nature that has ever been made; and, dying in the calm certainty of a life well spent, he would of course leave his notebooks to the Royal Society. Would the Royal Society thank him for the treasure of a lifetime of observation? It would not. The Royal Society would treat his notebooks exactly as the English bishops have treated Joanna Southcott's box. It would refuse to open them at all, because it would know, without looking, that the notebooks contain only a jumble of disorderly and meaningless items.

The reason is that as Paul Weiss has said: 'The primary aim of research must not just be more facts and more facts, but more facts of strategic value'.

It is interesting that Albert Einstein had once remarked that one of the great gifts that he had was a good nose e.g. know what to look for, identify the essentials and see the importance and relevance of an approach.

Phenomenological Studies of the Cosmic Radiation

I refer in this paper to the phenomenological approach, because the field in which I worked was that of cosmic radiation, which provides a very good example concerning the power of phenomenological methods, as well as of the interdisciplinary approach in making discoveries. I started my research career in this field at the University of Bristol in England in 1949 working with Prof. C.F. Powell.

In the Nobel lecture that Prof. Powell gave in 1950, he spoke about the discoveries that studies in this field had led to. In particular, he gave an interesting quotation from the 1928 edition of the 'Conduction of Electricity in Gases', by J.J. Thomson and G.P. Thomson:

> It would be one of the romances of science if these obscure and prosaic minute leakages of electricity from well-insulated bodies should be the means by which the most fundamental problems in the evolution of the cosmos came to be investigated.

As time has gone on, this has increasingly proved to be the case.

It was in 1785 that Henry Coulomb noted that a metal sphere that was suspended by an insulator thread invariably lost the charge placed on it; this was more than what could be attributed to the loss of charge through the thread itself. Later, gold leaf electroscopes seemed to lose their charge at sea level, even though they were well-insulated electrically. C.T.R. Wilson, who invented the cloud chamber and was awarded a Nobel Prize for this work, connected this phenomenon with ionization of the surrounding air. This could have been due to radiation at ground level; this appeared possible after the discovery of radioactivity by Becquerel. There were unsuccessful efforts to reduce this by shielding the electroscopes. To explore this phenomenon further, in 1912, the Austrian scientist Victor Hess, flew instruments on balloons which he piloted himself. He observed that the phenomenon increased rapidly up to the highest altitude of around 6 kilometers attained by the balloons; and further, remained the same during the day and at night. He concluded that this was due to the incidence of a highly penetrating ionizing radiation from outside the earth, but not from the sun. Later this radiation was given its present name of cosmic radiation by Millikan.

Ever since this initial discovery, scientists studying cosmic radiation have significantly followed a phenomenological approach. They have studied the various components of the radiation, in terms of their nature as well as intensity, and variations of these as function of height in the atmosphere, depth below ground, across latitudes, in relation to events on the Sun etc; and garnered a remarkable amount of information.

We now know that cosmic radiation is like a very thin rain of particles impinging on top of the earth's atmosphere. They consist of neutrinos, electrons and protons, and charged nuclei of elements from helium up to uranium. Protons constitute the most abundant component. Their energies range over 15 orders of magnitude, up to energies far greater than any accelerator can, or is conceived to, produce on earth. Being charged particles, they are subject to the influence of magnetic fields. The earth's magnetic field, whilst enabling all the lower energy particles to pour in at the poles, prevents radiations with energies below a geo-magnetic cut-off coming in at other latitudes, with the highest cut-off being at the equator. The low energy component is also subject to variations caused by electro-magnetic conditions in the earth-sun vicinity. These radiations, in passing through the earth's atmosphere, collide with the air nuclei; and in these collisions are born new short-lived nuclear entities. In addition, there is a total breakdown of the heavy nuclei into their component fragments; as also the generation of showers and cascades.

The study of cosmic radiation gave birth to a whole new field of elementary particle physics, including discoveries relating to transient nuclear entities that live for very short time periods. The major discoveries in this field were of the mysterious muon and verification of the existence of the positron, the first example of antimatter that was theoretically predicted by Dirac. This was by Carl Anderson using a cloud chamber equipped with a magnetic field. The phenomenon was observed and then interpreted. They were not looked for, in the sense of conducting an experiment to verify the prediction of a theory. Then followed the observation (discovery) of electron positron pairs, as predicted by Dirac. Apart from these discoveries, after the Second World War, a whole new world of subatomic particles was discovered. These included the charged and neutral pions, and their decays, strange particles (concerning whose discoveries and modes of decay I had worked), excited states of the nucleon etc. These discoveries were made using nuclear emulsion techniques and cloud chambers. These discoveries gave an impetus to the building of larger and larger accelerators for elementary particle investigations. The field was dominated by cosmic ray observations for two decades, from the mid-thirties to the mid-fifties; it was then taken over by the accelerators. But the very high energy cosmic ray interactions, including those that produce the giant air showers, continue to remain the domain of cosmic ray observations.

In the preceding paragraphs it will be noted that I have also indicated the detector systems that were used for making the various discoveries. These detector systems did not exist in nature. The development of these, therefore, was not really a discovery, but an invention. But these inventions resulted in discoveries. And invariably, when a new technique was brought into play, nature yielded up some of its secrets.

Prior to the discovery of the cosmic ray window into the Universe, one only had the electro-magnetic window. This covered a narrow optical wavelength interval. All that we knew about the Universe then was due to the photons of visible light. Observations through this window were greatly enhanced through the invention of the spyglass or telescope by Galileo. In the mid-thirties, the discovery of radio waves from objects and regions in the Universe greatly expanded the electromagnetic window. Today, there are giant telescopes operating across a wide spectrum of photon wavelengths: from the radio waves, the infrared, visible, ultraviolet and X-ray bands; different types of information flow through these windows.

We are aware that there is a major force of gravitation that pervades the Universe and one would hope to be able to detect gravitational radiation – but this is yet to be observed. One became aware of phenomena that would give rise to neutrinos from space, but because of their extremely small cross-section for interaction, difficulties in observing them were anticipated. Only recently have the clear signals of neutrinos from the sun, as well as of galactic neutrinos been detected – these have called for the invention of the giant detector systems.

The particle radiation from the Universe, including the neutrinos, that the cosmic radiation represents is unique in many ways. The charged particles are subject to electromagnetic fields; their acceleration to the extraordinarily high energies seen is also the result of electromagnetic fields. They interact strongly with matter, this causes changes in their composition. The neutrinos come direct from the source, since they interact so weakly and are not subject to electromagnetic fields. What we can gather about the Universe from the particle radiation is, thus, very different from that derived from the electromagnetic radiation.

The field of cosmic radiation has thus provided windows into the Universe that were not otherwise available. In most cases, discoveries made in the field of cosmic radiation were not expected nor planned for. They were entirely observational in nature and, therefore, passive experimentation. The discoveries were, thus, of a phenomenological nature, aided by inventions of systems for better observation of the phenomena.

Often, the new detectors were invented and because of that new phenomena observed.

The field has also been one that represented interdisciplinarity – on the one hand relating to nuclear elementary particle physics and, on the other hand, to astrophysics and cosmology. And, today, in our efforts to understand the Universe, the very small and the very large are coming close together. All this started with the initial observation of the leakage of electricity from well-insulated bodies.

Deep Underground Cosmic Ray Experiments – A Personal Account

I would like to recount one example of a phenomenological approach with which I was closely associated, that opened up an entire new field. As mentioned earlier, cosmic ray physicists had made efforts to observe the radiation from a variety of aspects: in terms of its composition, energies, phenomena that they gave rise to, variation with altitude, latitude and depth. With regard to the last, namely, the depth variation of cosmic rays below ground, we had a group working at the Tata Institute of Fundamental Research, Bombay. It had initially measured how cosmic ray intensity changes as one goes below ground – in railway tunnels or shallow mines where there is some overburden of earth. However, in India, in the Kolar Gold fields, depths up to 2 miles below the earth's surface are available. An experiment was, therefore, mounted to establish the depth-intensity curve up to the greatest feasible depth. When the detectors (Geiger-Muller Counter Telescopes, 3 square meters in area) were taken to the greatest depth of 9,600 feet below ground, no counts were observed in a period of 60 days. This clearly indicated that the intensity of the atmospheric muon component had been attenuated to a very significant extent. We then calculated what might be expected from neutrinos produced in the atmosphere, traversing the earth, and interacting to produce muons in the rock surrounding the detector. For neutrinos in the GeV energy range, it appeared that the number of events of neutrino-induced muons passing through the detector would be adequate for observation. We immediately mounted an experiment with a suitable neutrino telescope; and were the first to observe the interaction of atmospheric neutrinos.

A similar experiment was done, somewhat around the same time, by Fred Reines and his collaborators in the South African ERPM Gold Mines. Reines was later awarded a Nobel Prize for his work on neutrinos, particularly the first observation of low energy reactor neutrinos along with his collaborator, Clyde Cowan; until this discovery, the neutrino was essentially a theoretical artefact created by Pauli. A major experiment on low energy neutrinos originating from the Sun was carried out by Ray Davis which earned him a Nobel Prize.

It is all of this work, starting in the mid 1960s which has given birth to the field of neutrino astrophysics. Many Russian and Japanese scientists have contributed to ideas and experiments in this field.

We have seen in this part of the paper, how observations carefully made of phenomena, have led to the opening up of wholly new fields: radioactivity and nuclear physics, antibiotics, and cosmic rays (the story of which has been described in some detail). Many similar examples can be alluded to.

One may well ask what is the future of such an approach for making discoveries. One has seen in one case almost a repeat of the phenomenological approach in the same field. Karl Jansky opened up the field of radio-astronomy, and decades later Penzias and Wilson discovered cosmic microwave background radiation; in both the cases these constituted careful observations relating to radio signals from space. Astronomy and astrophysics, in particular, is full of fairy tale stories of discoveries such as pulsars, black holes, dark matter and so on.

It is clear that the phenomenological approach will continue to pay rich dividends. Perhaps only the scale of the effort and the nature of detector systems will perhaps undergo major changes. But as long as one is dealing with individuals with a sense of curiosity, who can discern the important from the unimportant, and have a good nose as Einstein remarked, discoveries will continue to be made.

Conclusion

One has seen extraordinary minds over the past few hundred years, such as Darwin, Faraday, Galileo, Gauss, Newton, Pasteur, Leonardo da Vinci; and over the last century, Bohr, Crick and Watson, the Curies, Dirac, Edison, Einstein, Fermi, Pauling, Poincare, Rutherford, Thomson, and many more. The windows they opened, and the profound understanding of nature that their original thinking led to, has undoubtedly been the base for so much that has followed in the areas of application with a succession of rapid and great innovations. The pioneers of nuclear energy and of the space programmes, of flight and the jet engine, of computer science and information technology, of molecular biology and biotechnology, of the many areas of chemistry and new materials, and of the underpinning mathematical sciences, are all too numerous to be named; comparisons are always difficult; and yet each one of these represents wholly new areas of human progress. Any one of these could have been picked up, to describe in greater detail, aspects relating to discovery, creativity and innovation.

What is important is that there was a fertile soil in society which could lead to the flowering of these extraordinary minds and the areas that they launched. This is the civilizational and cultural framework of a scientific society. But for it to be truly in the public good, it must be complemented by creativity in the many other domains which ultimately make a complete human being and define values needed for human happiness.

REFERENCES

1. Ghosh, Sumit, 'Creative Thinking', *IEEE Circuits and Devices Magazine* (Jan/Feb. 2004), pp. 52-62.
2. Hadamard, Jacques, *Essay on Psychology of Invention in the Mathematical Field*, Dover Publications Inc. (Princeton University Press, 1945).
3. Kanigel, Robert, *The Man Who Knew Infinity*: *A Life of The Genius Ramanujan*, Scribners (London, UK, 1991).
4. Menon, M.G.K., 'The Characteristics and Promotion of Innovation', *OECD Proceedings – Social Sciences and Innovation* (2001) chapter 7, pp. 77-88 (Meeting held in Tokyo in 2000).
5. Poincaré, Henri, *Mathematical Creation in the Foundations of Science*, trans. G. Bruce Halstead (The Science Press, New York, 1913), (Lecture delivered at the Société de Psychologie in Paris).

THE ANTHROPOCENE: THE CURRENT HUMAN-DOMINATED GEOLOGICAL ERA

PAUL JOSEF CRUTZEN

Figure 1 shows me more than 70 years ago in the lap of my grandmother. I have changed a lot, but so has much on our planet. Human population has increased threefold during my lifetime, reaching about six thousand million, with the largest rise, 1.8% per year, after the Second World War. As shown in the partial listing of Table 1, many human activities impact on earth's environment, often surpassing nature with ecolog-

Figure 1.

TABLE 1.

Anthropocene

- During the past 3 centuries human population has increased tenfold to 6000 million
- Cattle population increased to 1400 million (one cow/family)
- Urbanisation has increased tenfold in the past century; almost half of the people live in cities and megacities
- Industrial output increased 40 times during the past century; energy use 16 times
- Almost 50 % of the land surface has been transformed by human action
- Water use increased 9 fold during the past century to almost 1000 m^3 per capita; 65 % for irrigation, 25 % industry, ~10 % households
- Fish catch increased 40 times
- The release of SO_2 (160 Tg/year) by coal and oil burning is at least twice the sum of all natural emissions; over land the increase has been 7 fold, causing acid rain, health effects and poor visibility due to aerosols
- More nitrogen is now fixed and used as fertilizers in agricultural than is produced by natural processes on land.
- Releases of NO to the atmosphere from fossil fuel and biomass burning is larger than its natural inputs, causing high ozone levels over extensive regions of the globe
- Several climatically important "greenhouse gases" have substantially increased in the atmosphere, eg. CO_2 by 30%, CH_4 by more than 100%.
- The production of chlorofluorocarbon gases has caused major changes in the chemistry of the stratosphere, leading to ozone loss, which has been particularly large in the spring over the Antarctic continent, the so-called ozone hole.

ical, atmospheric chemical and climatic consequences. Under the pressures of growing human populations, their need for food, water, and improved living conditions, land use changes have been substantial. For instance, during the period 1990-1997 the global annual rate of deforestation has been 0.5%/year with a maximum of 0.9%/year in South-East Asia. Human activities have especially accelerated after the Second World War, several examples of which are shown in Figures 2a and 2b. Starting

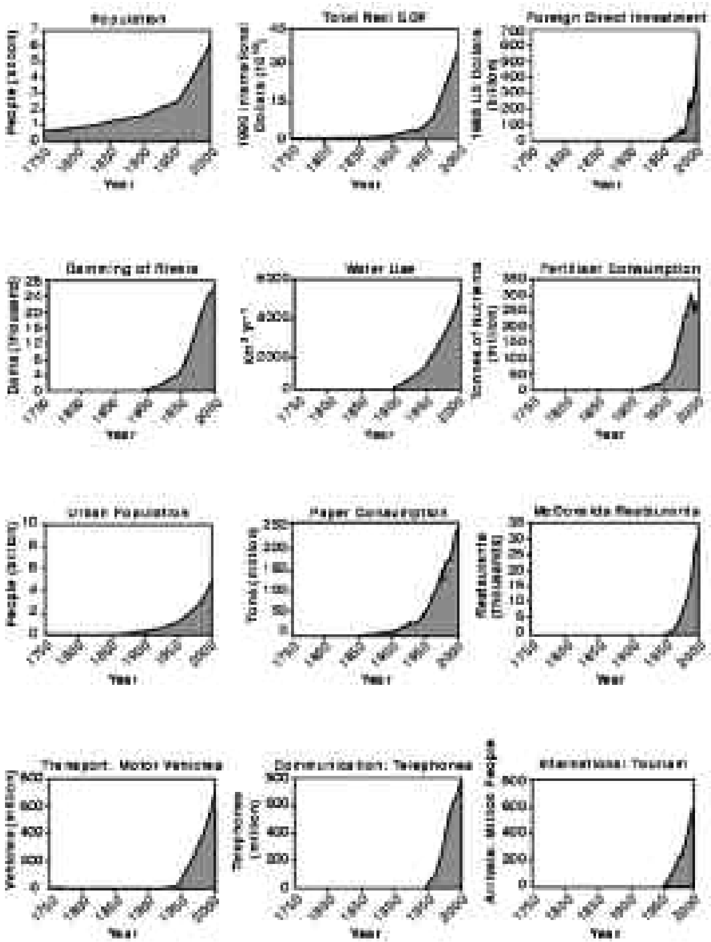


Figure 2a.

especially by the end of the 18th century, the growing disturbance of earth's natural systems by human activities created a new geological era, which I have dubbed the 'Anthropocene'.

In this paper, we will discuss the consequences of human activities on the earth's atmosphere and climate during the 'Anthropocene'. Figure 3 shows the probably most cited diagrams on the consequences of human activities on the atmosphere. The upper curve shows the rise in atmospheric carbon dioxide concentrations since the end of the 1950s, as meas-

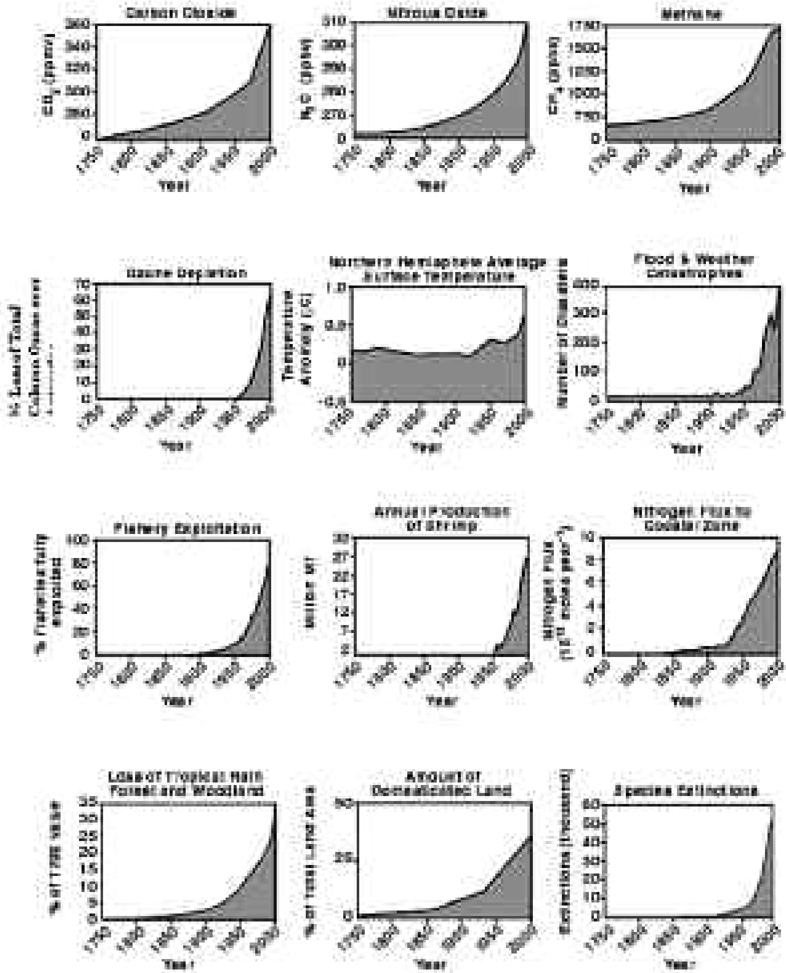


Figure 2b.

ured at the Mauna Loa observatory on the island of Hawaii. Superimposed on a long-term upward trend are seasonal variations with lowest values in the northern hemisphere during spring and summer, reflecting uptake of atmospheric CO_2 by vegetation during the growing season, and highest values in fall and winter due to the release of CO_2 from rotting vegetation. The lower two panels in Figure 3 show the 'ozone hole', the drastic loss of ozone which now regularly occurs in the stratosphere over the Antarctic continent

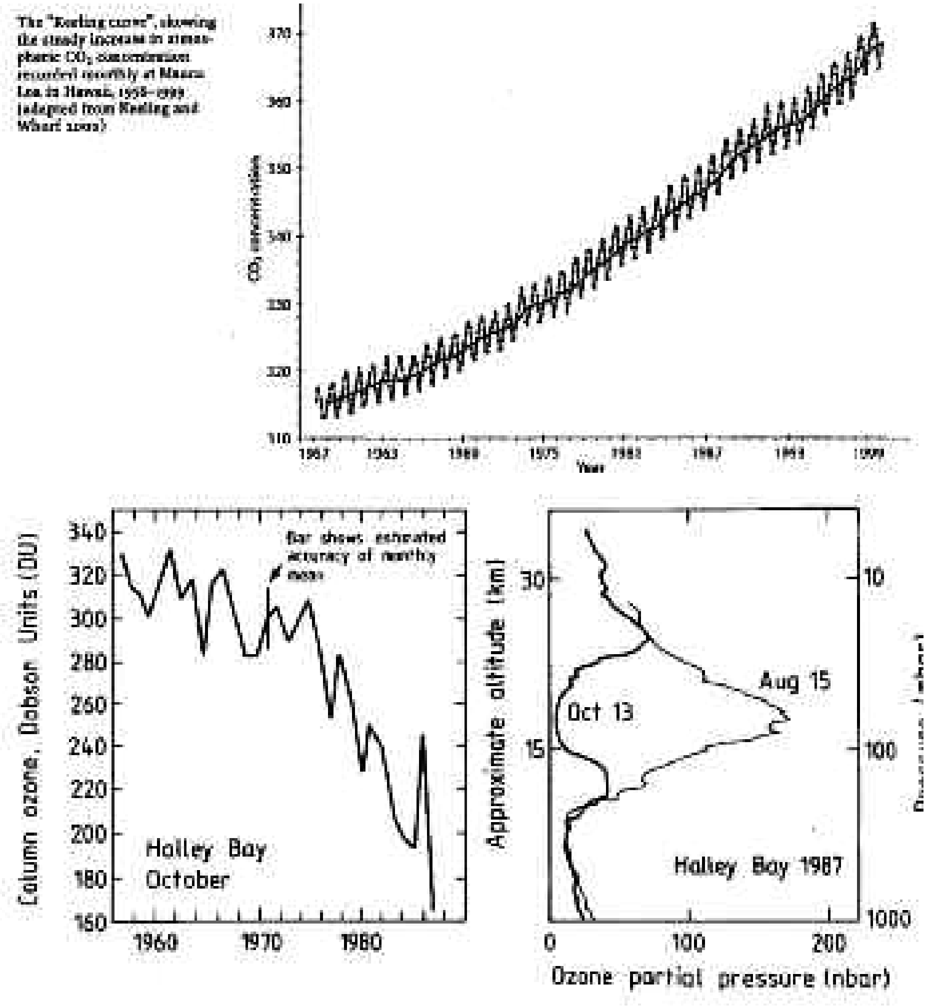


Figure 3.

in springtime. The total ozone column depicted in the lower left, shows a total ozone loss since the 1960-1970s by a factor of two in October. After two months ozone is completely destroyed between about 15 and 22 km altitude, exactly where naturally maximum ozone concentrations are found (right, lower panel). The ozone hole came as a total surprise. It is due to a number of originally unexpected, positive feedbacks, which start with the development of very cold temperatures, colder than about 195 K, during the south polar winter, causing formation of ice particles and production of very reactive, industrial chlorofluorocarbon-derived, ozone-destroying chlorine atoms and chlorine oxide molecules by catalytic reactions on the particle surfaces in the presence of sunlight during late winter/early spring. Although emissions of chlorofluorocarbon (CFCs) gases no longer take place in the industrial world, because of the longevity of the CFCs, the ozone hole will not heal before near the middle of this century. The ozone hole could only develop by a series of positive (reinforcing) feedbacks, something to remember when we later discuss the human influence on earth's climate system.

In Figure 4 (see page 292) we depict estimated changes in radiative forcings relative to pre-anthropocene, A.D. 1750, conditions. The radiative climate forcing due to the growth in the atmospheric greenhouse gases – CO_2, CH_4, N_2O, the CFCs, and tropospheric ozone – add up to about 2.7 W/m^2. This heating term is well-known. Solar variability constitutes a radiative forcing by up to about 0.3 W/m^2, about 10% of the anthropogenic greenhouse warming. Land use changes through urbanization and deforestation, and albedo changes by ice sheet and glacier melting, also affect earth's radiation budget by a similar amount, but do not constitute dominating factors. The most important, but also most uncertain, negative radiative forcings are related to the presence of aerosol particles (sulfates and organics) and clouds in the atmosphere. These terms are very uncertain, thus the large error bars. Except for black carbon (soot) containing aerosol, in general particles exert a negative forcing on earth's radiation budget due to their propensity to scatter a fraction of the incoming solar radiation back to space. Another important negative forcing is related to what is called the aerosols' indirect effect, which is caused by the increased reflectivity of sunlight by clouds due to enhanced number concentrations of aerosol, which serve as cloud condensation nuclei (CCN). As a consequence, for the same liquid water content, but with larger CCN number concentrations, the polluted clouds will consist of a greater number of smaller droplets, increasing the surface of cloud droplets, in turn leading to enhanced backscattering of sunlight to space by the clouds. Although it may sound contradictory, dirty clouds are brighter than clean clouds. Furthermore, under such conditions, rain formation, which depends on coalescence of cloud droplets to create bigger precipitating drops, also decreases, thus diminishing cloud precipitation efficiency, producing longer-lived clouds and increasing cloudiness.

A tentative energy budget derived by the author and Prof. V. Ramanathan, also a member of the Pontifical Academy, for present atmospheric conditions compared to pre-anthropocene conditions is shown in Figure 5. As mentioned before, greenhouse radiative forcing is 2.7 W/m^2. Uptake of heat by the oceans over the period 1955 to 1998 on average was 0.3 W/m^2. Enhanced upward flux of infrared radiation at the 'top of the atmosphere' due to the increase by 0.6 K in global average surface temperature cools the planet by about 1 W/m^2, assuming unaltered relative humidity in the atmosphere, which is a reasonable, but not totally proven, assumption. This implies an energy surplus by 1.4 W/m^2, which earth can only have got rid of by increasing its albedo via enhanced backscattering of solar radiation by clouds and anthropogenic aerosols. This is about half of the greenhouse infrared radiative heating, indicating the possibility of a doubling in fully realized greenhouse gas forcing, if the atmosphere would become free from anthropogenic aerosol. Future reductions in air pollution, which by themselves are desirable because of the negative health effects of atmospheric particles, may, therefore, double climate warming by the greenhouse gases, constituting a major policy dilemma.

It is here very informative to compare the infrared greenhouse gas radiative forcing of 2.7 W/m^2 with the direct heat release by humanity's

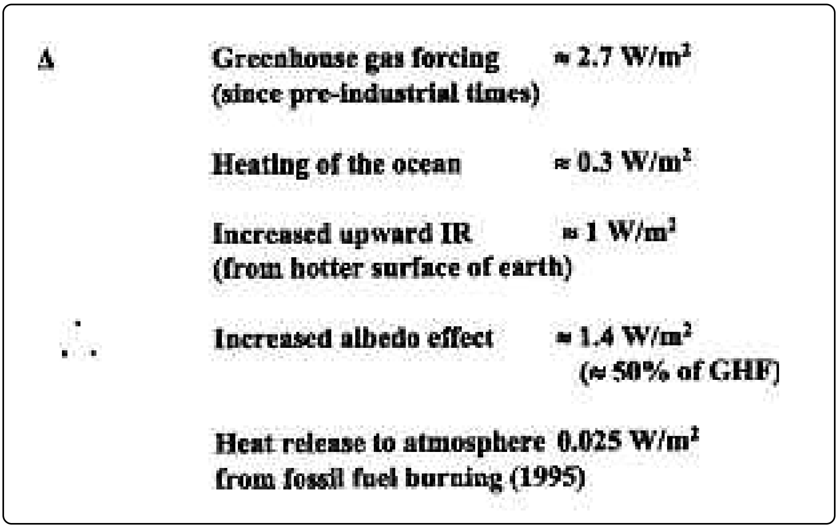


Figure 5.

energy use, which is only 0.025 W/m^2. Thus production of the greenhouse gases, and not the direct heating of the atmosphere by energy production, is the problem. A major effort to develop alternative energy sources other than burning of fossil fuels is the only way to prevent excessive climate warming. Carbon dioxide is not the only factor. Together, other greenhouse gases have provided almost as much greenhouse warming as CO_2. Reductions in their emissions may be more easily possible, for instance CH_4 from fossil fuel production, cattle holdings, and rice fields, N_2O from the application of nitrogen fertilizers in agriculture, and NO from automobiles, which results in a reduction of tropospheric ozone.

As shown in Figure 6, many complex and uncertain processes, with positive, as well as negative, feedbacks determine earth's climate. In accordance with what is expected from the growth in anthropogenic greenhouse gases, as shown in Figure 7, global mean temperatures on earth have increased since 1860 when direct temperature readings became possible in many parts of the world. The temperature rise since 1970 can currently only be explained by the anthropogenic greenhouse effect. A significant part in the rise in global temperatures over the period 1910 to 1940, however, reflects contributions from factors other than greenhouse gas emissions, such as heat release from the ocean, internal climate variability, or increase in solar

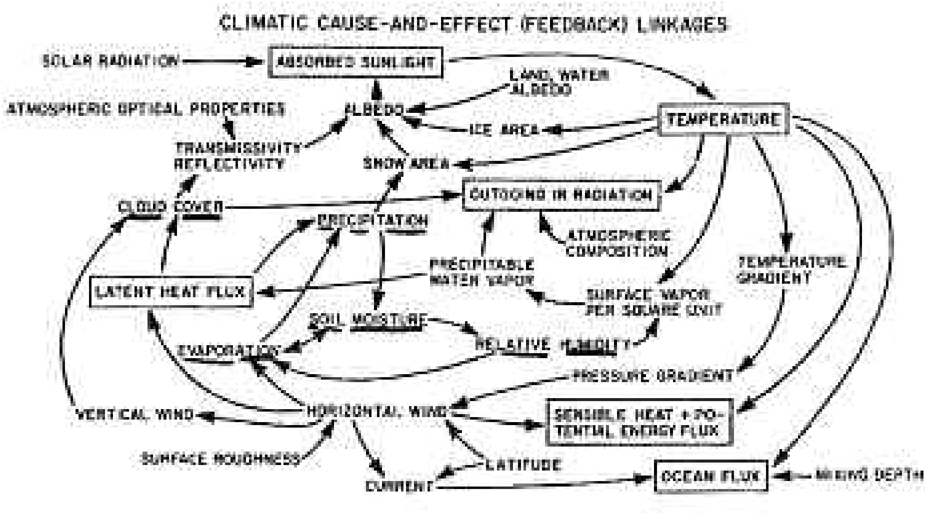


Figure 6.

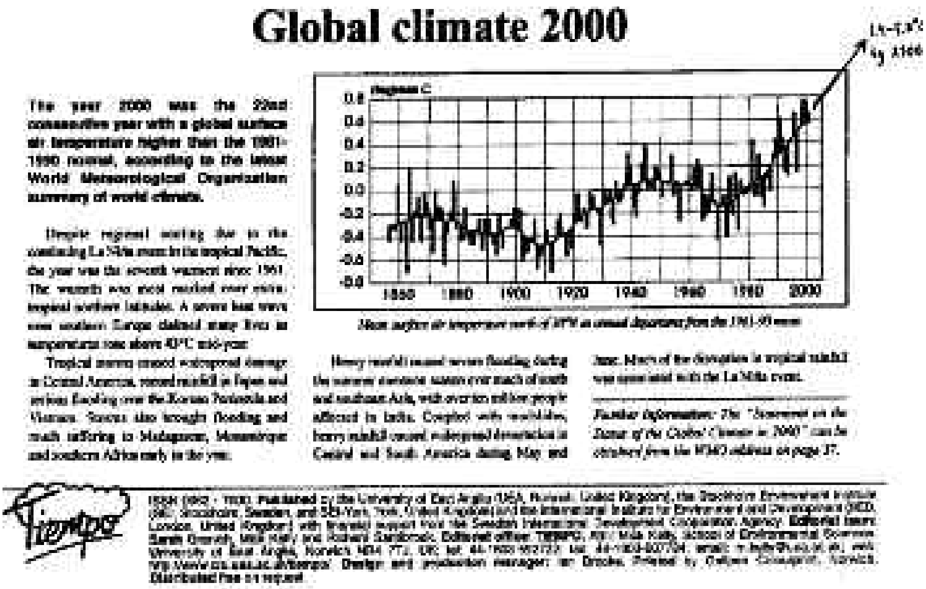


Figure 7.

irradiation. If the latter were the main factor in climate warming, as some claim, it would imply a very high climate sensitivity $dT_s/dQ=0.4°C/(0.3\ W/m^2)=1.3°C\ m^2/W$, where dT_s is the increase in global average surface temperature and dQ the increase in radiative forcing in Watts per m^2. With such a high climate sensitivity, the expected equilibrium surface temperature rise caused by the current greenhouse gas and aerosol forcings, according to the simple budget derived earlier, would be about 1.3 x 1.4 = 1.8°C, much larger than the observed 0.6°C (see Figure 7).

Earlier, the Intergovernmental Panel on Climate Change (the IPCC) had projected global average surface temperatures increases by 1.4-5.8 K over the current century, the large range being due to uncertainties, in particular in the modelling of the hydrological cycle (clouds, precipitation) and future scenarios of greenhouse gas and particle emissions to the atmosphere. The IPCC report of 2001 stated: 'There is new and stronger evidence that most of the warming observed over the last 50 years is attributable to human activities'. The warming would be accompanied by sea level rises of about 10-90 cm during the present century. Thus, uncertainties are large, not to speak about potential surprises as in the ozone hole case, briefly discussed before.

To prevent major global warming, drastic reductions in greenhouse gas emissions are needed. To stabilize the atmospheric concentrations of global CO_2 and N_2O, anthropogenic emission reductions by some 60% are called for, a very difficult, if not impossible, task in a world in which the majority of nations are still in a state of underdevelopment, as shown in Figure 8 (see page 292), which depicts the current per capita use of carbon in metric tons per year. Unfortunately, as shown in Figure 9, Nature does

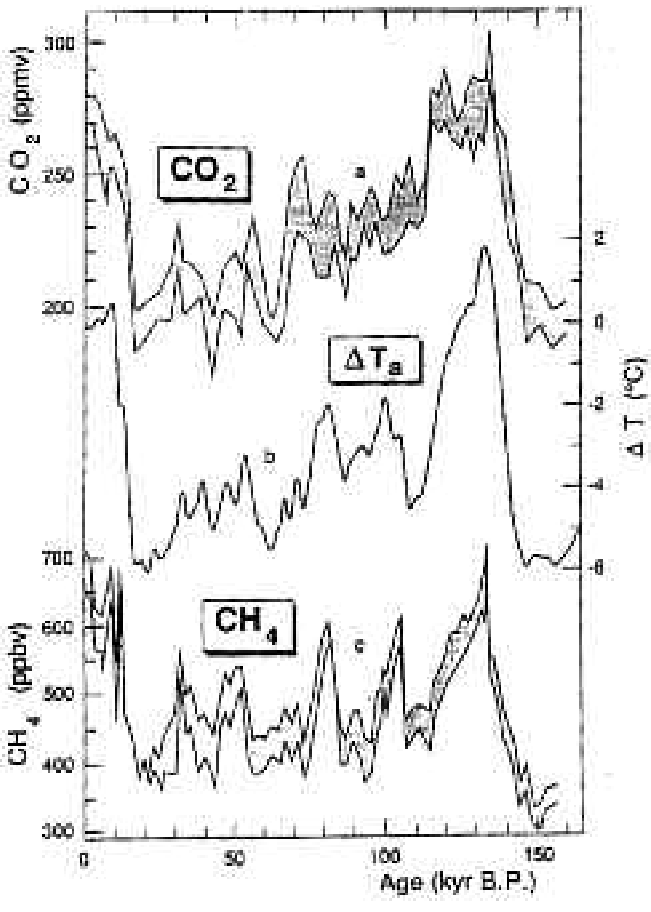


Figure 9.

not come to the rescue. During the past million years (the figure only shows the past 150,000 years) at any time when temperatures, derived from ice core data, were high, also the atmospheric concentrations of CO_2 and CH_4 were highest, thereby actually strengthening climate warming which is initially triggered by the behaviour of earth's orbital parameters. A major uncertainty is the stability of earth's climate system with all its complications. Especially sensitive may be the climates of the polar regions where largest climate warming is observed and also calculated by models. Questions have especially been raised about the stability of the thermohaline ocean circulations and the North Atlantic gulfstream which transports heat from lower latitudes to Europe.

Although earth's climate may be most vulnerable at high latitudes, the greatest future impact of human activities on earth's climate may well be in the tropics, especially in rapidly industrialising Asia. Heavy air pollution now covers large parts of the tropics and subtropics with high loadings of aerosol particles, containing relatively large amounts of black carbon, which produces the so-called Asian Brown Cloud. This is clearly visible from space, showing an extensive haze plume over India and flowing southward over the Indian Ocean, towards the Maldives.

Besides soil dust, the aerosol consists of a mixture of sulfates, organics, ammonia, and black carbon (see Figure 10, page 293), produced by fossil fuel and biomass burning, and by agricultural activities. Especially because of the strong absorption of solar radiation by the black carbon fraction of the aerosol, major disturbances in the radiative budget of the atmosphere over south, southeast and east Asia occur during the dry season. As a result, substantial absorption of solar radiation takes place in the atmosphere which can reach values of some 30 W/m^2 over the Asian continent during the dry season, which lasts from November to May. This is accompanied by a similar decrease of the solar energy absorption at the earth's surface. These numbers are about ten times larger than the global anthropogenic energy disturbance at the 'top of the atmosphere'. As a consequence of the enhanced energy uptake in the atmosphere and decrease at the earth's surface, the atmosphere will warm and the surface will cool, leading to a reduction in water evaporation at the earth's surface, reduced cloudiness and less rainfall, and slow-down of the hydrological cycle, which may also affect regions downwind of the polluted regions. This Atmospheric Brown Cloud (ABC) is most pronounced over the Asian continent, but is also present over Africa, and South America. The formation and persistence of the so-called brown clouds are favoured by the long dry season which hinders the removal of the aerosol particles by precipitation. Because of the strongly growing economic development of Asia, the ABC phenomenon will become an increasingly major factor in anthropogenic regional climate change and rainfall characteristics, especially over Asia. But there remains the dilemma: combating air pollution by reducing aerosol emissions, may very well lead to a warmer climate.

Finally, to finish where I began, Figure 11 (see page 293) shows me in 1999 together with my then two-year- old grandson Jan Oliver. And I ask myself how the earth will be like during his lifetime which will cover most of the twenty-first century. What is clear is that in order to keep the earth habitable, major restrictions are needed in the use of earth's resources, below ground, at the earth's surface and in the atmosphere. Mankind has a long way to go when it comes to a wise use of natural resources.

REFERENCES

Crutzen, P.J., 'Geology of Mankind: The Anthropocene', *Nature*, 415, 23 (2002).

Crutzen, P.J., and V. Ramanathan, 'Atmospheric Chemistry and Climate in the Anthropocene', chapter 4 in *Earth System Analysis for Sustainability* (MIT Press, 2003).

Crutzen, P.J., and V. Ramanathan, 'The Parasol Effect on Climate', *Science*, 302, 1679 (2003).

Hofman, D.J., *et al.*, 'Ozone profile measurements at McMurdo Station, Antarctica, during the Spring of 1987', *J. Geophys. Res.*, 94, 16527 (1989).

Houghton, J.T., *et al.* (eds.), *Climate Change 2001; The Scientific Basis* (Cambridge University Press, 2001).

Keeling, C.D., and T.P. Whorf, 'Atmospheric CO_2 records from sites in the SIO sampling network', *Trends, a Compendium of Data on Global Change*, Oak Ridge National Laboratory, USA DOE (2000).

Lelieveld, J., *et al.*, 'The Indian Ocean Experiment', *Science*, 291, 1031 (2001).

Levitus, S., Antonov, J.I., and J. Wang, 'Anthropogenic Warming of Earth's Climate System', *Science*, 292, 267 (2001).

McNeil, J.R., *Something New Under the Sun* (Norton, New York, 2000).

Petit, J.R., *et al.*, 'Climate and Atmospheric History of the Past 420,000 Years from the Vostok Ice Core, Antarctica', *Nature*, 399, 429 (1999).

Rahmstorf, S., 'Thermohaline Circulation: The Current Climate', *Nature*, 421, 699 (2003).

Ramanathan, V., Crutzen, P.J., Kiehl, J.T., and D. Rosenfeld, 'Aerosols, Climate and the Hydrological Cycle', *Science*, 294, 2119 (2001a).

Ramanthan, V., *et al.*, 'The Indian Ocean Experiment', *J. Geophys. Res.*, 106, 28371 (2001b).

Sateesh, S.K., and V. Ramanthan, 'Large Differences in Tropical Aerosol Forcing at the Top of the Atmosphere and Surface', *Nature*, 405, 60 (2000).

Steffen, W., *et al.* (eds.), *Global Change and the Earth System* (Springer, 2004).

'Global Climate 2000', *Tiempo* (University of East Anglia, 2001).

DISCOVERING THE PATHWAY FOR COOLING THE BRAIN

ROBERT J. WHITE

Introduction

The brain, in spite of extensive research, remains one of the greatest enigmas of human existence. Admittedly, this living organ of only 1500 grams of weight, composed of trillions of cells and fibers formed in a jelly-like matrix, is responsible for all we know as human beings and understand of ourselves and about the universe itself. Since the beginning of human existence the entirety of what man has accomplished, both in a positive and negative way since the beginning of human history, is a result of the activity of this most complex and unique object in the known world. It is within the physical structure of its living fabric that specialized functions that characterize the human mind such as cognition, memory, and intelligence, to name only a few, are all anatomically located in the brain substance. Thus, it can be appropriately argued the theme of this symposium that will deal with the process of creativity and discovery, obviously, must be based on the functioning of the human brain/mind.

This particular presentation will examine the investigative activities undertaken to discover pathways to efficiently cool the brain without causing any measurable alterations in its higher functions. The overall reason for this fundamental form of research is the overwhelming sensitivity of brain tissue at normal temperature (37°C) to any interruption of its intrinsic blood flow. It has been scientifically established that only 5 to 6 minutes of total circulatory arrest to the brain under normal thermal conditions is required to permanently limit or destroy its physiological and psychological performances.

Sensitivity of Cerebral Tissue to Ischemia and Ways to Protect It

In recent years, one of the issues raised regarding lengthening this 5 to 6 minute interval was the possibility of developing methods that would provide protection to the cerebral organ during extended periods of circulatory arrest. In actuality, there are two basic problems here that require solutions. First, to find a technique that will offer significant protection during or following cerebral blood flow cessation. Second, to develop a biotechnology that can accomplish this purpose without damaging the brain or its performance capabilities. Obviously, the possibility of a suitable pharmacological agent that would offer these advantages has been extensively explored but, to date, without major success with the possible exception of barbitural compounds which appeared to offer minimal protection.

For some years now, the studies of environmental experience with survival following human cold immersion and hibernating animals has strongly suggested that by purposely alternating the thermal state of the brain, one might confer on the organ an extended time period of survival following the elimination of its somatic circulation but, that still leaves the question: would cooling of the brain be safe? And, could a technique be devised that would effectively cool the organ without being injurious to the cerebrum? Thus, at this point, in our experimental studies we had to discover the effects (be they good or bad) of temperature reduction on the central nervous system. In addition, we would have to create an engineering system that would be highly successful in altering the thermal state of the brain without producing any measurable damage to its intrinsic tissue composition which would cause functional degradation.

The Isolated Brain Preparation and Hypothermia

Early on, we felt it would be appropriate to develop an isolated brain model in order to document the pure physiological, biochemical, biophysical and reological changes resulting exclusively from vascular cooling of the brain. In this series of experiments, the entire brain including distal brain stem was surgically removed from the cranial vaults of a series of adult subhuman primates (Rhesus monkeys) and maintained in a viable state employing for circulatory purposes either a cross-circulation vascular arrangement utilizing a large Rhesus monkey (Fig. 1) or employing a specially designed miniaturized extracorporeal perfusion circuit (see Fig. 2, page 294). This would be equipped with a small oxygenator unit, a special-

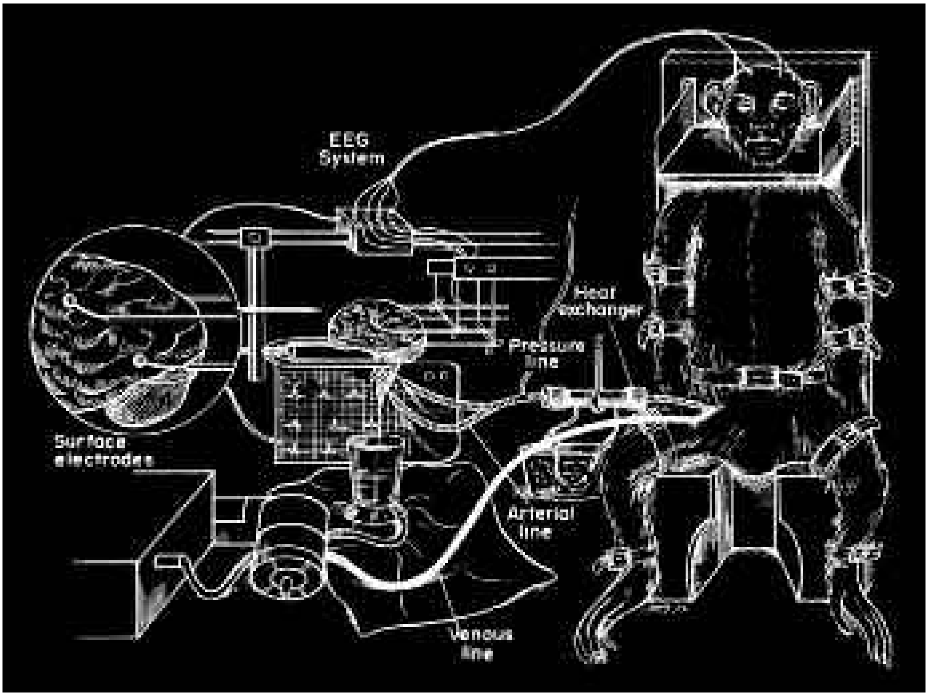


Figure 1. Isolated monkey brain maintained in a viable state employing cross circulation with large donor (note heat exchanger for brain cooling).

ly designed mini-mechanical compression pump, and a miniaturized heat exchanger, all linked together with a plastic tubing arrangement. Individual brains were instrumented (see Fig. 3, page 295) with surface (cortex) and, on occasion, depth electrode systems as well as needle thermometers for continuous intracerebral temperature measurements. From time to time, the isolated brains were submerged in a tank of artificial cerebral spinal fluid (CSF) and often with ventricular draining systems.

With these extremely delicate subhuman living brain modules under conditions of total isolation (available for the first time in medical history), a series of experiments were designed to provide answers to the crucial questions outlined above. In the case of the cross-circulation supported isolated brains, they responded much as a transplanted organ with no evidence of rejection and with normal physiological and biochemical profiles following repeated cooling, arrest of circulation, and re-warming.

As a result of these studies, the significant depression of the electrical and metabolic activities of the total brain at various low temperatures was documented for the first time.

When the isolated brain preparations were supported exclusively by a fully independent extracorporeal perfusion circuit they actually performed very well early on, however, after some time (8 to 10 hours), the brain began to display evidences of failure with decrease of electrical and metabolic activity as well as the development of cerebral edema and subpial bleeding. To some degree, this could be improved by replacing the perfusate (diluted monkey blood with Dextran) entirely with a fresh perfusate. These experiments conclusively demonstrated that, for the first time, the brain could be maintained as a viable isolated organ with a totally mechanized circulation. To significantly extend organ survival, the equipment and perfusate required improvement and this could be appropriately accomplished without too much difficulty.

Educated Monkeys and Differential Brain Cooling

Finally, we asked ourselves what would be the best experimental design to truly answer the remaining question in these experiments; have there been any significant changes in the behavior of these monkeys that resulted from the cooling technique and global brain ischemia under conditions of profound hypothermia? To address this problem, we felt strongly that we would have to employ highly trained subhuman primates to be completely assured that in the intact animal there were truly no substitutive changes in their mental capacities. As a consequence, a group of monkeys were intensively trained for 6 months to a full year employing cognitive tests (Wisconsin testing device) before they were subjected to differential brain cooling and cerebral circulatory arrest.

With simplification of the perfusion circuitry (elimination of the oxygenator and mechanical pump) to an arterial circulatory system utilizing the animal's own heart to pump the blood through the system, it now contained only a miniaturized heat-exchanger to alter the blood's temperature as it entered the brain. The vascular arrangement required not only surgical exposure of the carotid and vertebral arteries in the monkey's neck, but frequently the femoral artery in the groin was also surgically exposed to provide an alternative site for arterial outflow. As a result of this arterial arrangement, blood flow into the brain could be limited to the carotid arteries and at the time of circulatory arrest all four arteries (carotid and vertebral) were occluded to provide global brain ischemia. Following periods of purposeful interruption of brain perfusion (1/2 hr. to 1 hr.,), the animal's brain was allowed to be re-warmed by the somatic circulation itself. Following a day or two of post-operative recovery, these subhuman primates were submitted for re-testing and demonstrated no evidence of intellectual decline. Thus, with the successful demonstration of cognitive performance in these highly educated animals after profound cooling and total ischemia of the brain, it was now believed to be time to examine the possibilities of creating a similar technology for the human being.

The Human Brain and its Thermal Relationship

As a result of its intracranial location, the brain is superbly protected against thermal variations in the environment The brain is literally immersed and floating in fluid (thus, its normal weight of 1500 grams is reduced to approximately 47 grams in accordance with Archimedes principle). In turn, it is surrounded by its bony skull, whose inner surface is lined by a non-elastic, thin, tensel strong membrane – the dura. Finally, the tissues of the scalp, including muscles, skin, and fat offer excellent insulation to preserve appropriate intracranial temperatures and prevent the development of wide variations within the thermal environment of the brain. As a consequence, cooling the brain via convection methods would appear to be extremely difficult unless, of course, the entire surface of the body was also cooled. This has been done on numerous occasions including to date, but has been found to be laborious, inefficient, and time consuming.

Thus, the problem before us was obvious; how to reduce the brain's temperature safely, rapidly, and efficiently. The solution, as far as we were concerned, would be fashioned after the technique developed in the monkey and would require using the body's own vascular system which, of course, was primarily responsible for homeostatically maintaining the thermal relationship of the entire body. We were also committed to examining the issue of brain hypothermia in an isolated state. In other words, we were determined to cool only the brain while the rest of the body's organs were kept at, or near, their normal temperatures. Obviously, to successfully accomplish our mission we would have to create a modified extracorporeal unit, similar in design to the circuit utilized in the subhuman primate, which would allow us to exclusively perfuse only the brain in situ while the somatic circulation remained intact and responsible for the rest of the body (see Fig. 4, page 296).

THE DEVELOPMENT OF DIFFERENTIAL COOLING OF THE HUMAN BRAIN

We were well aware of the great success that extracorporeal engineering systems (the heart-lung machine) had accomplished for cardiac surgery. Often, they were additionally equipped with instrumentation that offered the capability of rapidly reducing brain temperature to 10° or below, especially in infants undergoing congenital heart surgery. We were also knowledgeable in regard to the neurological complications which frequently developed with the use of extracorporeal circulation during surgery especially when body (and brain) cooling was utilized These unfortunate cerebral events were often, in the beginning, believed to result from the employment of profound cooling. In time, however, it was established to be an embolitic phenomena caused by the interaction of blood and the surfaces of the tubing and machinery. In order to minimize this possibility, we eliminated the pumping element and the oxygenator from our cerebral perfusion unit. As a consequence, our artificial circulation became an arterial-to-arterial perfusion system, not requiring artificial propulsion. In this vascular design, the individual's own heart provided the circulatory force to propel the blood through the system. Also, since the vascular circuit is totally arterized, no oxygenating equipment was necessary (see Fig. 5, page 297). While this has seen only limited employment in human brain surgery, there has been no evidence of neurological damage.

THE PATHWAY FOR BRAIN COOLING: DISCOVERY AND RESULTS

Thus, we come to the end of our discovery pathway which has led us to a safe, rapid and efficient technique for providing profound hypothermia of the brain to provide protection under conditions of staged total cerebral circulatory arrest. It is almost unimaginable to think that one can duplicate the same conditions in the human brain that were proven to be so successful in the subhuman primate brain, where deep hypothermic temperatures were produced followed by prolonged periods of total cerebral circulatory arrest without producing evidence of structural or functional damage, it has been achieved. However, the question remains; how does a biophysical substance which possesses memory and cognition retain and preserve these unique properties as well as many other special functions following this dramatic reological and thermal distortion? If one looks at its tissue morphology under the electron microscope and examines its intrinsic electrical activity and its mental capabilities, we have no scientific explanation as to why the physical structure of this living material can be reduced to an extremely low temperature (5°C.), where all evidence of neuroelectricity has ceased and metabolism reduced to a minimum and, yet, when vascularly re-warmed to normothermia will demonstrate normal neuroelectrical, neurochemical and retain behavior-appropriate behavioral performance. What has been so conclusively achieved in animal studies should be applicable to the human central nervous system as well. We must not forget that in both animal and human investigations significant periods of total brain ischemia from 30 to 60 minutes (10 times the possible time frame for survival of the brain as a functioning organ at normothermia) were imposed with complete neurological and psychological recovery as a result of the brain being maintained in a deep hypothermic state.

In the End, the Human Brain Studied the Brain

One of the difficulties that arises when studying the brain/mind is that one is actually investigating the organ with itself – thus, to create pathways to discover and unravel the mysteries and activities of the brain/mind provides an interesting paradox – could the brain fool the brain itself? Or, could it purposely, or even unwillingly, mislead the human brain/mind in terms of the creative and discovery process? Nevertheless, we must acknowledge that *all* creative activity and the origin and design of all pathways of discovery exclusively originate in the human brain.

THE IMPACT OF MICROBIAL GENETICS ON THE DEVELOPMENT OF GENOMICS AND BIOTECHNOLOGY

WERNER ARBER

Introduction

In the last 60 years, research in the life sciences has uncovered a wealth of information on biological functions. This has mainly become possible by introducing new research strategies including the experimental exploration of biologically active molecules and their interactions, in using among other means quantitative measurements of relevant parameters, improved imaging techniques, biophysical, biochemical and molecular biological approaches as well as bioinformatic tools. In this article, special attention will be paid to developments that have their origin in microbial genetics. Within this context we shall highlight a few particular discoveries in which the author had been personally involved.

Roots and Early Development of Genetics, Evolutionary Biology and Nucleic Acids Biochemistry

Classical genetics goes back to Gregor Mendel who described in 1866 that phenotypic traits of peas become transferred into the progeny and that plants with different traits could give rise to recombinants upon sexual reproduction. This scientific approach started to flourish in the first half of the 20th century.

Contemporarily with Mendel's findings, Charles Darwin published in 1859 the theory of natural selection representing the major root of evolutionary biology which can explain the wide natural biodiversity that forms a pillar of ecology.

It is around 1940 that scientists built a bridge between the hitherto independently developed fields of genetics and evolutionary biology. This so-called modern evolutionary synthesis explained that genetic variants are at the basis of altered phenotypic traits and that they also form together with their parents the substrate for natural selection and thus the evolutionary development. At that time, histological evidence suggested that genes reside in chromosomes, but their chemical nature remained unknown.

A publication made in 1874 by Friedrich Miescher is at the root of nucleic acids biochemistry, which in 1953 culminated in the description by Watson and Crick (1953) of the double helical structure of DNA insuring a high degree of fidelity upon the replication of the filamentous DNA molecules.

The Early Days of Microbial Genetics

Bacteria are haploid, unicellular microorganisms that reproduce by cell division. For a long time they were thought to have no sexuality. It is probably for this reason that geneticists did not pay attention to bacteria (Zuckerman and Lederberg, 1986). This changed only in the early 1940s when microbial mutations were isolated and when particular mixtures of cells with different identified traits yielded recombinants (Lederberg and Tatum, 1946; Lederberg, 1947). At the same time Avery, MacLeod and McCarty (1944) could show that DNA, rather than protein, is the carrier of genetic information. In the crucial experiment they mixed highly purified, hence protein free, pneumococcal DNA with living cells of another strain of Pneumococcus bacteria. Under these conditions some of the latter cells acquired traits characteristic for the strain that had been the source of the DNA. It took about ten years until the biologists recognized the importance of this transformation experiment and the conclusion regarding the carrier of genetic information. The likely reason for this delay is seen in the fact that chromosomes are composed of DNA and of proteins. The higher degree of complexity of proteins, as compared to that of nucleic acids, had stimulated the scientists of that time to assume that the postulated complex nature of a gene could best be explained by proteins. The breakthrough to the acceptance of the conclusions of Avery *et al.* (1944) was greatly helped by the knowledge of the linear structural features of DNA molecules. This knowledge opened new avenues to explore the genetic code and the embedding of genes into the long DNA filaments.

Different Natural Means Provide Opportunities for Horizontal Transfer of Genetic Information Between Different Strains of Bacteria

As we have already seen, free DNA molecules liberated by a donor bacterium may sometimes be taken up in transformation by a genetically distinct acceptor strain (Avery *et al.*, 1944). In contrast, the process first explored by Lederberg (1947) is based on a kind of sexual mating between two bacterial cells. In this process of conjugation, donor DNA becomes linearly transferred into the acceptor cell through a bridge built between the mating partners. A third natural possibility of horizontal gene transfer depends on a viral particle as a vector for a segment of the donor genome. This process is called transduction and it was first described by Zinder and Lederberg (1952). It will catch our specific attention below.

In any process of transfer of genetic information from a donor to an acceptor cell, the transferred DNA must become firmly associated with the acceptor genome, should it get inherited into the progeny of the acceptor cell. Microbial genetics has described different natural ways to accomplish this requirement by recombinational processes or by maintaining the acquired DNA molecule as an autonomously replicating unit, a so-called plasmid.

Cointegration of a Viral Genome with the Bacterial Host Genome

It had been seen in the early days of microbial genetics that some bacterial viruses could strongly associate with their host bacteria for a more or less extended time. In this situation the host survives the infection and it propagates. Once in a while the genetic information of the virus can become active in one of the progeny cells and produce viral progeny particles. The host cell will thereby die by lysis. The underlying phenomenon is known as lysogeny. Its explanation goes back to studies by Lwoff and Gutmann (1950). Further explorations later revealed that some viruses integrate their genome temporarily into the host genome, while other viral genomes are rather maintained as plasmids.

When a cointegrated, endogenous virus, also called provirus, becomes activated again, its genome becomes excised from its chromosomal location. This is brought about by the same site-specific recombination enzyme that is also responsible for the cointegration in the establishment of lysogeny.

Study of Proviral Mutants

Most of the viral genes remain silent, unexpressed in the proviral state. A provirus may thus accumulate spontaneous mutations in the course of time. Some of these mutations will inactivate genes that are essential for viral reproduction. The study of such mutations was the topic of my PhD thesis. Some of the mutants analyzed were unable to produce intact viral particles upon reactivation of virus reproduction. The deficient structural elements (such as empty heads, filled heads, tails, association of empty heads with tails of the bacterial virus) could be quantitatively analyzed in the electron microscope (Arber and Kellenberger, 1958). This allowed us at least in some cases to identify the specific gene function that had been hit by the mutation.

In this electron microscopical study we also included a viral derivative which transduced the genetic information of bacterial origin encoding the fermentation of the sugar galactose (Morse, Lederberg and Lederberg, 1956). In its proviral form this derivative was unable to produce any viral structures visible in the electron microscope. The only way out of this surprising situation was to undertake genetic studies. These investigations revealed that a relatively important segment of the viral genome was fully absent. Obviously, it must have been substituted by a segment of the host genome carrying the genes responsible for galactose fermentation together with a few other genes (Arber, Kellenberger and Weigle, 1957, 1960; Campbell, 1962). As we will see below, this hybrid structure associating host genes with a part of the viral genome became in the early 1970s a model for gene vectors serving in recombinant DNA technology.

Horizontal Gene Transfer Encounters Several Natural Barriers

In the early years of microbial genetics, it became rapidly obvious that the efficiency of horizontal gene transfer varied widely in function of the genetic and evolutionary relatedness of the bacterial strains involved. Barriers act against DNA acquisition at various steps of horizontal gene transfer. First, the surface of the acceptor cells must be compatible with the needs for a successful uptake of the donor DNA. Second, enzymes of restriction-modification systems can distinguish between foreign DNA and the cell’s own DNA, as we will more clearly explain below. Third, the transferred DNA has to become firmly associated with the acceptor genome in order to become effective in the progeny of the hybrid. And, finally, the newly acquired genetic functions must be compatible with the functional harmony of the entire genome in order to withstand the pressure of natural selection.

The Discovery of Restriction Enzymes

In their work with bacterial viruses involving more than one bacterial host strain, several investigators observed in the 1950s that the involved virus grew very inefficiently upon a change of the host bacteria (see Arber, 1965b). In general, the few progeny viruses obtained could, however, efficiently re-infect the new host. But often, once adapted to the new host, the viruses did not any longer infect efficiently their previous host bacteria. This phenomenon was called host-controlled modification. The adaptation to a new host was called modification, and the inefficiency of infection upon the change of the host was called restriction. Since restriction was observed also upon periodic back and forth changes between a pair of two distinct hosts, the scientists correctly argued that modification could not be explained by a genetic mutation. Therefore, many scientists thought that modification was brought about by some host protein which became associated with the viral particles.

In 1960 I became unexpectedly confronted with the molecular basis of host-controlled modification. As a postdoctoral investigator at the University of Geneva I got engaged in a project to study the effects of different types of radiation, such as ultraviolet light, X-rays and radioactive decay, on living organisms. This research on biohazards of radiations was carried out in the context of a Swiss national program of research in view of the peaceful use of atomic energy. Our intention was to carry out our studies with different strains of *Escherichia coli* bacteria and with bacteriophage λ. In the course of preparing several *E. coli* strains which should also serve as hosts for growing the bacterial virus λ we encountered the phenomenon of host-controlled restriction. Driven by the curiosity to understand the molecular basis of this phenomenon we undertook a series of one cycle growth experiments. These revealed that against the general assumption it was the phage DNA rather than a host protein which was the subject of modification (Arber and Dussoix, 1962) and which was also the target for restriction. Indeed, restricted phage DNA became rapidly degraded upon infection of a restricting host and this explained the high inefficiency of the infection (Dussoix and Arber, 1962). Interestingly, in other experiments it had been shown that phage DNA which had suffered radiation damage became also degraded upon infection, even in non-restricting hosts (Kellenberger, Arber and Kellenberger, 1959). For a while we wondered if the DNA degradation observed in the different types of experiments had the same roots. This stimulated us to follow the different situations in parallel. In addition, this argument served to justify our experimental investigation of host-controlled modification in the framework of the research project on biohazards of radiations, although this sideline had not originally been foreseen in the project.

Modification is an Epigenetic Phenomenon and Consists in Sequence-Specific Methylation of Nucleotides

Since DNA restriction and modification appeared to become of high prospective interest (Arber, 1965b) we decided to concentrate our upcoming research to the further exploration of this phenomenon. These studies resulted in the insight that the molecular basis of modification, the adaptation of a virus to grow efficiently on a given host strain, resided in the methylation of a nucleotide that is imbedded into a specific sequence of nucleotides of a length of four to about ten base pairs (Arber, 1965a; Arber and Linn, 1969; Smith, Arber and Kuhnlein, 1972). The attached methyl group affects neither the normal base pairing nor the correct expression of the concerned genes. Modification is thus an epigenetic phenomenon.

The Search for Restriction and Modification Enzymes

On the basis of the described findings it was postulated that bacteria usually possess one or more restriction and modification systems serving as a kind of immune defense against foreign DNA entering into the cell. Restriction enzymes were postulated to act as nucleases destructing foreign DNA. The cell's own DNA was postulated to be insensitive to this restriction cleavage because of its proper modification, the methylation in the strain-specific DNA recognition sequences. Within a relatively short time, this interpretation was confirmed by the isolation of restriction endonucleases and modification methylases and by the study of their *in vitro* activities (Meselson and Yuan, 1968; Linn and Arber, 1968; Smith and Wilcox, 1970; Kuhnlein and Arber, 1972).

In comparative studies of the activities of purified restriction enzymes it was confirmed that restriction cleavage becomes indeed activated on specific recognition sequences on the DNA as long as these sites carry no strain-specific methylation. We now know that some restriction enzymes (type II enzymes) cleave their substrate DNA molecules precisely at the recognition site (Roberts *et al.*, 2003), while some other restriction enzymes (type I) translocate the DNA after recognition and eventually cleave it at a more or less random location (Murray, 2000). Since the 1970s the type II enzymes widely serve as tools in genetic analysis and engineering.

In Vitro Recombinant DNA Techniques

A major difficulty in the attempts to study genetic functions at the molecular level remained still around 1970 the tremendous size of the filamentous DNA molecules carried in the chromosomes. In the search for means to sort out DNA fragments of handsome size, appropriate for sequence analysis and functional analysis, the scientists became aware of the naturally observed possibility of a covalent association of a given DNA segment with an autonomously replicating vector DNA molecule. We have already encountered this phenomenon with some bacterial viruses and it had also been shown to occur with conjugative plasmids (Jacob and Adelberg, 1959; Adelberg and Pittard, 1965). Experiments carried out to produce such hybrid DNA molecules *in vitro* in using a bacteriophage λ derivative as a vector were successful (Jackson, Symons and Berg, 1972; Lobban and Kaiser, 1973). This not only allowed the investigators to sort out a specific DNA segment from its genomic location, it also enabled them to amplify the sorted-out segment in order to obtain enough well purified material to carry out structural and functional analyses.

As soon as type II restriction enzymes became available genetic research benefited from their reproducible DNA cleavage function producing manageable DNA fragments. This enabled the researchers to establish physical genome maps (restriction cleavage maps). Specific DNA segments could be sorted out and used to produce *in vitro* recombinant DNA molecules (Cohen *et al.*, 1973).

By these developments, based on scientific knowledge on natural processes of specific interactions of bacterial enzymes with DNA molecules, molecular genetic studies became possible for every kind of living organism. Just a few years later, still another microbial enzyme, a thermo-resistant DNA polymerase was at the basis of the introduction of the polymerase chain reaction. This PCR reaction enables the researchers to highly amplify specific DNA segments at their natural location under the only condition that short flanking sequences are already known (Saiki *et al.*, 1988).

Search for Nucleotide Sequences and Functions of DNA

Still in the 1970s, chemically based strategies were developed to determine the nucleotide sequences of selected and amplified DNA segments (Sanger, Nicklen and Coulson, 1977; Maxam and Gilbert, 1977). Once the DNA sequences became known, one could envisage to undertake functional studies on selected open reading frames as well as on elements controlling gene expression or maintenance functions of the DNA molecules. For this purpose strategies of local site-directed mutagenesis were developed (Shortle, Di Maio and Nathans, 1981; Smith, 1985). This enables the researchers to compare the phenotype of the wild type form of a gene with those of designed mutants. Alternatively, the deletion of a DNA segment or other kinds of DNA rearrangements by methods of genetic engineering can also serve for site-specific mutagenesis. These approaches unravel quite often, although not always, the biological function encoded by the gene or other genetic element in question.

Comparison of Research Strategies Used in Classical Genetics and in Molecular, Reverse Genetics

Investigations in classical genetics depend on the availability of mutants. These can have a spontaneous origin or they can be induced by a treatment with a mutagen. The mutant is recognized by an altered phenotype which becomes transmitted into the progeny. The phenotypic changes can give a reliable hint to the specific function affected by the mutation. Genetic crosses between independently isolated mutants serve to establish genetic maps, and specific genetic information can be localized on these maps. Note, however, that this approach of classical genetics does not depend on any knowledge on the chemical nature of the carrier of genetic information. In classical genetics the concept of the gene remains an abstract notion, without physical entity.

In contrast, most investigations of the new molecular genetics start with isolated DNA molecules with the aim to identify their biological functions. This research goes from the carrier of genetic information to the functions, while research in classical genetics, as we have just seen, goes from functions to a genetic map. In view of this strategic difference of the research approaches, molecular genetics is sometimes also called reverse genetics. In this strategy a manageable, well purified DNA fragment is sequenced, open reading frames and potential control signals are identified, site-directed mutations are then placed on strategic spots of the DNA under study, the mutated DNA segment is introduced instead of its wild type form into the cell under study, alterations in the phenotypes as compared to the wild type condition are looked for and this can, at least sometimes, allow one to conclude on specific biological functions of the DNA segment in question. This strategy is generally applicable to the genomes of most living organisms, at least with some appropriate modifications. It represents the essential basis for genomics and to some degree also for proteomics.

Note that the definition of a mutation differs in molecular genetics (changed nucleotide sequence) from that used in classical genetics (phenotypic change). Not all changes in a given nucleotide sequence will become manifested by a changed phenotype, while an inheritable change in a phenotype is always caused by a change in the nucleotide sequence.

Impact of Functional Genomics on the Development of Biotechnology

Both the molecular genetic research strategies and the thereby acquired knowledge offer wide novel possibilities for biotechnological applications. Generally speaking, biotechnology takes advantage of biological functions for the benefit of mankind and increasingly also of its environment. Such applications may, for example, specifically relate to an improvement of human, animal or plant health, to nutritional security, to agricultural production or to environmental remediation.

Specific knowledge on particular biological functions as a result of investigations in functional genomics can offer ample possibilities to make use of these functions in biotechnology. Thereby, methods of molecular genetics such as site-directed mutagenesis can serve for improvements of the functions in question, both with regard to their quality and quantity. Most importantly, the strategies of molecular genetics render it possible to transfer a given specific genetic information into other organisms that may sometimes be unrelated to the original source of the biological function in question. This can, for example, be of high relevance for the biotechnological production of a gene product to serve as a medical drug. Recall that in classical biotechnology, that has been practiced for many centuries, one has to use the organisms as they are found in nature. At most, one can try to improve a function or the yield of a product by breeding techniques and by random mutagenesis. Still today this does often not include a thorough molecular genetic analysis of the resulting hybrids and mutants, respectively. In contrast, genetic modifications carried out with modern molecular genetic strategies usually include a careful analysis of the modified organisms both at the genetic and functional levels.

Conjectural Risks of Genetic Engineering

At a very early time in the application of *in vitro* recombinant DNA techniques, the involved scientists themselves raised the question of possible biohazards related to some of their experiments. In order to debate these questions an International Conference was held in February 1975 in Asilomar (Berg *et al.*, 1975). In brief, possible risks of genetic engineering may become manifested in a short-term or in a long-term. Pathogenicity, toxicity, allergenic effects and other harmful or undesirable effects can be counted among the short term risks. These can be carefully studied experimentally within a reasonable time before any of the genetically modified organisms are approved for biotechnological applications. In order to protect preventively the health of researchers and more generally that of the human population, appropriate guidelines were drawn up and these require that according to a scientifically based classification of a given risk, the research has to be carried out under precautions that are worldwide in use in medical diagnosis of pathogenic microorganisms.

The prediction and identification of long-term risks of genetic engineering is a more difficult task than that of the evaluation of short-term risks. Long-term risks may sometimes have an impact on the course of biological evolution, particularly with genetically modified organisms that are deliberately released into the environment, as it is for example the case for agricultural crops. As a matter of fact, the production and release of genetically modified organisms represents a human contribution to biological evolution. An important requirement to responsibly evaluate any long-term evolutionary risks associated with intentional genetic alterations is a good knowledge of the natural process of biological evolution. In view of this consideration, I decided at the Asilomar Conference held in 1975 to concentrate my own future research on studies of the process of biological evolution at the molecular level.

Three Qualitatively Distinct Molecular Strategies Contribute to the Spontaneous Formation of Genetic Variants

Fortunately, a considerable amount of data on molecular mechanisms of spontaneous genetic variation was already available in the 1970s, mainly from microbial genetics. Many more data were subsequently obtained from specifically designed research projects. For these reasons, it is now possible to draw reliable conclusions regarding the mechanisms and strategies that contribute under natural conditions to genetic variation. At previous occasions I have reported to our Academy on this progress (Arber, 1997, 2002, 2003a). I will therefore just briefly summarize here the main facts and conclusions.

Several different specific molecular mechanisms, rather than a single mechanism, contribute to the formation of genetic variants. These mechanisms can be classified into three general strategies that possess different qualities with regard to their contribution to genetic evolution.

One strategy brings about small local changes in the sequences of the genome, such as a nucleotide substitution, the deletion or the insertion of one or a few nucleotides, or a scrambling of a few nucleotides. Some of these changes, in particular the substitution of a single nucleotide, can valuably contribute to the evolutionary improvement of existing biological functions. To make this point clear, it should be kept in mind that by far not each nucleotide substitution will result in a functional improvement. Rather, it is natural selection that favors rare spontaneous beneficial variants according to the rules of Neodarwinism. Local sequence changes can be brought about by replication infidelities involving often natural structural flexibilities (tautomerism) or chemical instabilities of the nucleotides, as well as by the action of chemical and some physical mutagens. In many of these cases nascent mutations are rapidly repaired by appropriate enzyme systems. For larger genomes the absence of efficient repair is detrimental for the organism.

A second strategy for the generation of genetic variants is a rearrangement of DNA segments within the genome. This DNA reshuffling depends in general on activities of recombination enzymes such as for homologous recombination, for so-called site-specific recombination and for transposition of mobile genetic elements. These processes can yield a duplication and higher amplification of a DNA segment, the deletion of a DNA segment, the inversion of a DNA segment, the translocation of a DNA segment and, as is widely known for diploid organisms, hybrid chromosomes with genes from the two parents. Some of these reshuffling processes can bring about novel gene fusions as well as the fusion of a given open reading frame with an alternative expression control signal. Again, rare favorable rearrangement products providing functional benefits will be favored by natural selection. More often, however, a DNA rearrangement will reduce the functional harmony of the genome and thus cause a selective disadvantage.

The third strategy of generating genetic variants is DNA acquisition by horizontal transfer of genetic information from a donor organism into an acceptor organism. This phenomenon is best studied with bacteria since it is at the basis of bacterial genetics. This involves bacterial conjugation, virus-mediated transduction and transformation by free DNA molecules as transfer processes. Depending on the evolutionary relatedness of the donor and acceptor strains, horizontal gene transfer can give rise either to conversion (the substitution of a segment of genetic information by a different, but still homologous DNA sequence) or to the acquisition of genetic information that was hitherto not present in the acceptor genome. Again, it will depend on natural selection if the resulting hybrid will be favored or not in the long-term. For the concerned acceptor organism, the successful acquisition of foreign genetic information can represent a rapid and efficient functional innovation. DNA acquisition can be seen as a sharing in successful developments made by others.

Comparison of Genetic Alterations Obtained by Genetic Engineering with Those Occurring Spontaneously

Genetic engineering uses the same three strategies of genetic variation that serve in the natural world for the purpose of biological evolution. Genetic engineering may indeed involve a local change of nucleotide sequences, it may bring about a rearrangement of genomic sequences or it may consist in the acquisition of a segment of foreign genetic information. Similarities between the natural and the intended genetic changes are also seen with regard to the size of DNA sequences involved in these processes

(Arber, 2002). In this view, similar conjectural biohazards may be expected from genetic engineering and from natural genetic variation and, as a matter of fact, from classical breeding strategies.

However, these similar processes will of course generally not yield identical products in view of the tremendous number of possible unique genomic sequences. Therefore, absolutely precise predictions cannot be made. From these considerations one may deduce that a careful, responsible handling and long-term control of any organisms that had deliberately been genetically modified by human intervention is justified. This relates both to products of genetic engineering and to those obtained by classical breeding strategies. Particular attention should be paid to organisms into which genetic information from a genetically unrelated donor organism had been inserted, because of a lack of solid knowledge on the range and the probability of successful horizontal gene transfer under natural conditions. In this context, it is relevant to recall that deliberate mass production, as it applies to many plants of agricultural use independently of their origin, favors just by statistical means their occasional involvement in evolutionary processes.

The Theory of Molecular Evolution

Besides its practical relevance for the evaluation of conjectural risks of genetic engineering, a profound knowledge of molecular mechanisms that serve in the natural world for the generation of genetic variations represents a basic contribution to a deeper understanding of biological evolution. What has in fact become possible in the last few decades is a second evolutionary synthesis, integrating molecular genetics and Neodarwinism to become a theory of molecular evolution (Arber, 2003b; 2004). This represents an expansion of the Darwinian theory of biological evolution to the level of molecular processes, particularly those involved in genetic variation, in reproductive isolation and eventually also in natural selection.

From the short description that we have given for the three strategies for genetic variation it is obvious that in most of these molecular mechanisms specific enzymes are involved. Genetic studies with microorganisms have shown that many of these enzymes are not essential for the normal clonal propagation of bacteria from generation to generation. This is, for example, the case for transposition of mobile genetic elements or for site-specific DNA inversion. However, these processes are of obvious relevance in the occasional production of genetic variants. The involved enzymes are the products of genes. In view of their functional relevance for biological evolution we call these genetic determinants evolution genes. The products of some of the evolution genes are actively involved in the production of genetic variants, as we have seen, these are in fact variation generators. The products of other evolution genes have the task to keep the frequencies of genetic variation low and tolerable for a long-term maintenance of the given kinds of organisms. These enzymes can serve for example in the repair of nascent mutations, or for the restriction of foreign DNA upon horizontal gene transfer.

The theory of molecular evolution postulates that the generation of genetic variations not only depends on activities of evolution genes. Rather, it assumes that a series of non-genetic elements play also their specific roles. This represents a making use of intrinsic properties of matter such as the tautomerism and chemical instability of nucleotides and various conformational flexibilities of biologically active molecules for the purpose of genetic variation. Other non-genetic elements influencing spontaneous mutagenesis are environmental mutagens and random encounter.

The Intrinsic Duality of the Genome

An interesting implication of the presence of the postulated evolution genes on the genome is a duality of the genomic information. We have to realize that not all the genes present on the genome exert their activities for the benefit of the individual organism in question. Other genes work for the benefit of the evolutionary development of the population. The evolution genes serve for a steady expansion of life, for the production and renewal of biodiversity. In contrast, the more classical housekeeping genes, accessory genes of use by all individuals under particular life conditions and developmental genes serve each individual for the fulfillment of its life. Note that the products of some genes can serve for both of these objectives and act for purposes of the individuals as well as of biological evolution.

Philosophical, World View Aspects of the Theory of Molecular Evolution

What has been outlined here on the basis of recently acquired scientific knowledge may have wide relevance for our worldview. We can generally conclude that natural reality takes active care of biological evolution, as it also takes care of individual physical lives. Mutations should not be considered as errors or as caused by accidents. Rather, intrinsic properties of matter together with activities of evolution genes are at their origin. Different specific molecular mechanisms, different natural strategies, contribute in specific ways to the process of biological evolution.

The genomic duality can offer an unexpected, at least partial, solution to the theodicean question. Variation generating evolution genes exert occasionally their mutagenic activity in a particular individual of a population. We have learned that new genetic variants are generally only rarely favorable, beneficial under the living conditions encountered by the concerned individual. More frequent are either neutral or unfavorable mutations. These might perhaps be favorable under other living conditions. In this regard, the evolutionary progress resembles a trial and error process with relatively few winners. Therefore, individuals having suffered an unfavorable mutation can be considered to have become victims of the natural process of biological evolution. Under the assumption that biological evolution is brought about by a divine intention and as a consequence of the genomic duality, both with regard to the presence of evolution genes and with regard to their variation generator activities, one can see a possible answer to the question of theodicy in the juxtaposition of physically good and physically evil in the overall genetic activities, deserving both the requirements of individuals and of evolving populations.

Conformity Between Traditional Wisdom and Scientific Knowledge on Biological Evolution

Consider the narration of creation as it is given in the Genesis as a testimony of traditional wisdom. Scientific theories and firmly established scientific knowledge are sometimes considered to antagonize traditional knowledge. This was also the case for the Darwinian theory of evolution. It is thus indicated to re-inspect the situation in a search for conformities between traditional wisdom and scientific views. According to the Genesis, God created our world stepwise. This can well correspond to the step-by-step progress of biological evolution. In addition, genetic variations must be at the basis of the well-distinguished personal characteristics of prophets and other descendants from the first human beings on our planet. These characteristics are quite specifically described in the Genesis, indicating that human beings were not considered as clones, they were rather seen as unique individuals. From a scientific point of view, this represents genetic diversity as a consequence of genetic variation. During creation God evaluated several times the quality of His work and He concluded that it was good. In today's scientific terms this includes the process of biological evolution as such, the generation of genetic variations and the genomic duality with its consequences that we have already described. Both, the health of the individual human beings and the prospective progress of biological evolution must correspond to God's will. From my laic point of view, I can see one of the missions of the son of God, Jesus Christ, to consist in teaching to the human society that it is a human duty to provide help for the suffering, and thus underprivileged people, by love and medical care. In the Christian faith, this can represent a responsible reaction to the theodicean problem that is linked to the process of continued creation anchored in biological evolution.

Concluding Remarks

After having been largely neglected by classical genetic research, microbial genetics, once initiated, has rapidly opened novel research strategies to identify DNA as the carrier of genetic information and then to investigate genetic information at the molecular level. This has given rise to molecular genetics that is applicable to all kinds of living organisms and that is now known as functional genomics. Some of these steps are here described in more detail, such as the identification of natural gene vectors and of restriction enzymes serving as valuable tools in molecular genetic research. It is also outlined how newly acquired knowledge on genetic functions can lead to fruitful biotechnological applications. In turn, such applications, particularly if they involve *in vitro* recombinant DNA techniques, raise questions of conjectural risks. Some of these risks relate to long-term evolutionary developments. Again, it is mainly on the basis of experimental data from microbial genetics and knowledge resulting from these experimental investigations that a theory of molecular evolution could be formulated. This theory postulates that spontaneous genetic variations are jointly caused by intrinsic properties of matter and by activities of evolution genes. This represents an expansion of the Neodarwinian theory to the level of molecular events involved in biological evolution. On the practical side, this forms a reliable basis for a responsible evaluation of long-term conjectural risks of genetic engineering. Besides this, the improved understanding of molecular processes involved in biological evolution has a strong impact on our world view, the fundament of the orientational knowledge that can serve the civil society to assume co-responsibility for practical applications of scientific knowledge for the benefit of mankind and its natural environment. Consistencies between religiously founded traditional wisdom and recently acquired scientific knowledge are discussed, as well as questions with regard to the simultaneous presence on the genome of genes acting for the benefit of the individual organism, and of evolution genes that insure the evolutionary progress of populations of organisms and thus a rich biodiversity.

REFERENCES

Adelberg, E.A. and Pittard, J. (1965), 'Chromosome transfer in bacterial conjugation', *Bacteriol. Rev*. 29, 161-172.

Arber, W. (1965a), 'Host specificity of DNA produced by *Escherichia coli*. V. The role of methionine in the production of host specificity', J. *Mol. Biol*. 11, 247-256.

Arber, W. (1965b), 'Host-controlled modification of bacteriophage', *Annu. Rev. Microbiology* 19, 365-378.

Arber, W. (1997), 'The influence of genetic and environmental factors on biological evolution', *Plenary Session on the Origin and Early Evolution of Life* (part I), The Pontifical Academy of Sciences, Commentarii, vol. IV, n. 3, pp. 81-100.

Arber, W. (2002), 'Molecular evolution: comparison of natural and engineered genetic variations', *The Challenges of Science*, The Pontifical Academy of Sciences, Scripta Varia 103, pp. 90-101.

Arber, W. (2003a), 'Cultural aspects of the theory of molecular evolution', *The Cultural Values of Science*, The Pontifical Academy of Sciences, Scripta Varia 105, pp. 45-58.

Arber, W. (2003b), 'Elements for a theory of molecular evolution', *Gene* 317, 3-11.

Arber, W. (2004), 'Genetic variation and molecular evolution', *Encyclopedia of Molecular Cell Biology and Molecular Medicine*, R.A. Meyers (ed.), Wiley-VCH, Weinheim, vol. 5, pp. 331-352.

Arber, W. and Dussoix, D. (1962), 'Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage λ', *J. Mol. Biol*. 5, 18-36.

Arber, W. and Kellenberger, G. (1958), 'Study of the properties of seven defective-lysogenic strains derived from *Escherichia coli* K12(λ)', *Virology* 5, 458-475.

Arber, W., Kellenberger, G. and Weigle, J. (1957), 'La défectuosité du phage Lambda transducteur', *Schweiz. Z. allg. Pathol. u. Bakteriol.* 20, 659-665.

Arber, W., Kellenberger, G. and Weigle, J. (1960), 'The defectiveness of Lambda transducing phage', *Papers on Bacterial Genetics*, E.A. Adelberg (ed.), Little, Brown & Co., Boston-Toronto, pp. 224-229.

Arber, W. and Linn, S. (1969), 'DNA modification and restriction', *Annu. Rev. Biochem*. 38, 467-500.

Avery, O.T., MacLeod, C.M. and McCarty, M. (1944), 'Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III', *J. Exp. Med*. 79, 137-158.

Berg, P., Baltimore, D., Brenner, S., Roblin, R.O. and Singer, M.F. (1975), 'Asilomar Conference on recombinant DNA molecules', *Science* 188, 991-994 and *Nature* 255, 442-444.

Campbell, A.M. (1962), 'Episomes', *Advanc. in Genetics* 11, 101-145.

Cohen, S.N., Chang, A.C.Y., Boyer, H.W. and Helling, R.B. (1973), 'Construction of biologically functional bacterial plasmids *in vitro*', *Proc. Natl. Acad. Sci. USA* 70, 3240-3244.

Dussoix, D. and Arber, W. (1962), 'Host specificity of DNA produced by *Escherichia coli*. II. Control over acceptance of DNA from infecting phage λ', *J. Mol. Biol*. 5, 37-49.

Jackson, D.A., Symons, R.H. and Berg, P. (1972), 'Biochemical method for inserting new genetic information into DNA of Simian virus 40: Circular SV40 DNA molecules containing Lambda phage genes and the galactose operon of *Escherichia coli*', *Proc. Natl. Acad. Sci. USA* 69, 2904-2909.

Jacob, F. and Adelberg, E.A. (1959), 'Transfert de caractëres gènètiques par incorporation au facteur sexuel d'*Escherichia coli*', *Comptes Rendus des Séances de l'Académie des Sciences* 249, 189-191.

Kellenberger, G., Arber, W. and Kellenberger, E. (1959), 'Eigenschaften UV-bestrahlter λ-Phagen', *Z. Naturforsch.* 14b, 615-629.

Kuhnlein, U. and Arber, W. (1972), 'Host specificity of DNA produced by *Escherichia coli*. XV. The role of nucleotide methylation in *in vitro* B-specific modification', *J. Mol. Biol.* 63, 9-19.

Lederberg, J. (1947), 'Gene recombination and linked segregation in *E. coli*', *Genetics* 32, 505-525.

Lederberg, J. and Tatum, E.L. (1946), 'Novel genotypes in mixed cultures of biochemical mutants of bacteria', *Cold Spring Harb. Symp. Quant. Biol.* 11, 113-114.

Linn, S. and Arber, W. (1968), 'Host specificity of DNA produced by *Escherichia coli*. X. *In vitro* restriction of phage fd replicative form', *Proc. Natl. Acad. Sci. USA* 59, 1300-1306.

Lobban, P.E. and Kaiser, A.D. (1973), 'Enzymatic end-to-end joining of DNA molecules', *J. Mol. Biol*. 78, 453-471.

Lwoff, A. and Gutmann, A. (1950), 'Recherches sur un Bacillus megathérium lysogëne', *Ann. Inst. Pasteur* 78, 711-739.

Maxam, A.M. and Gilbert, W. (1977), 'A new method for sequencing DNA', *Proc. Natl. Acad. Sci. US*A 74, 560-564.

Meselson, M. and Yuan, R. (1968), 'DNA restriction enzyme from *E. coli*', *Nature* 217, 1110-1114.

Morse, M.L., Lederberg, E.M. and Lederberg, J. (1956), 'Transduction in *Escherichia coli* K-12', *Genetics* 41, 142-156.

Murray, N.E. (2000), 'Type I restriction systems: sophisticated molecular machines', *Microbiol. Mol. Biol. Rev*. 64, 412-434.

Roberts, R.J., Vincze, T., Posfai, J. and Macelis, D. (2003), 'REBASE: restriction enzymes and methyltransferases', *Nucleic Acids Res*. 31, 418-420.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988), 'Primer-directed enzymatic amplifications of DNA with a thermostable DNA polymerase', *Science* 239, 487-491.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977), 'DNA sequencing with chain-terminating inhibitors', *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.

Shortle, D., DiMaio, D. and Nathans, D. (1981), 'Directed mutagenesis', *Annu. Rev. Genet*. 15, 265-294.

Smith, H.O. and Wilcox, K.W. (1970), 'A restriction enzyme from *Hemophilus influenzae*. I. Purification and general properties', *J. Mol. Biol.* 51, 379-391.

Smith, J.D., Arber, W. and Kuhnlein, U. (1972), 'Host specificity of DNA produced by *Escherichia coli*. XIV. The role of nucleotide methylation in *in vivo* B-specific modification', *J. Mol. Biol*. 63, 1-8.

Smith, M. (1985), '*In vitro* mutagenesis', *Annu. Rev. Genet*. 19, 423-462.

Watson, J.D. and Crick, F.H.C. (1953), 'Genetic implications of the structure of deoxyribonucleic acid', *Nature* 171, 964-969.

Zinder, N. and Lederberg, J. (1952), 'Genetic exchange in Salmonella', *J. Bacteriol*. 64, 679-699.

Zuckerman, H. and Lederberg, J. (1986), 'Postmature scientific discovery?', *Nature* 324, 629-631.

ENDOSYMBIOTIC BACTERIA ASSOCIATED WITH PLANT SEEDS AND BIRDS' EGGS

CRODOWALDO PAVAN

FOREWORD

During the last three years, in cooperation with a colleague of mine at the University of São Paulo – Brazil, we have been coordinating a research project on 'Nitrogen Fixing Bacteria in non-legume plants'.

Three labs were involved in the development of the project. My group in Lab 1 was responsible for finding, isolating and purifying the strains of bacteria collected in nature and agricultural plantations. The two other laboratories were respectively responsible for the taxonomical classification and the molecular characterization of the bacteria species collected. During the first two years, after the start of the cooperative project, our group collected and purified over 50 species of nitrogen fixing bacteria from different non-leguminous plants, along with several other bacteria that were used in the master theses of two students involved in the project. Unfortunately, the taxonomical identification and the molecular characterization of the collected nitrogen fixing bacteria, for which the two other labs were responsible, were not done for various reasons and this motivated me to enhance a habit I normally have which is to individually analyze, in addition to the basic problems of the main project, which I was part of, anything interesting that occurs as a sideline. This had already happened to me in the 1950s, when I was working with Prof. Theodosius Dobzhansky on 'Natural population of Drosophila in Brazil'. At that time I discovered *Rhyncosciara angelae*, an extraordinary insect not related to Drosophila, but excellent to cytogenetics and gene action studies. In that case the work on Drosophila was proceeding very successfully and later we published excellent papers, which was expected having Dobzhansky as the leader of the groups. But even so during the development of the main project I was looking for any other interesting thing that appeared in our field excursion. In one of them I discovered groups of larvae of *Rhyncosciara angelae* and with this exceptional insect I later published alone or with colleagues several papers which are among the most important in my scientific carrier.

In the research with fixing nitrogen bacteria the result was a failure, but the sideline observations resulted in an important discovery. I have discovered the existence of endosymbiotic bacteria living normally in great number inside seeds of plants and eggs of birds which I believe is a discovery important enough to be described here.

Introduction

The title of this article is: 'Endosymbiotic bacteria associated with plant seeds and birds' eggs' which is part of a more general project that we are organizing entitled: 'Endosymbiotic bacteria in plants and birds'.

Endosymbiotic bacteria are part of the body of the organism with which they are associated and are present not only in the seeds or eggs of the respective organisms.

In plants it is easier to follow the distribution of the endosymbiotic bacteria in the body of the host, as parts of the embrions in development, roots, stem, leaves and finally in the flowers which are responsible for the origin of new seeds.

Bacteria are the most ancient, the structurally simplest, and the most abundant of organisms, and are the only ones characterized by a prokaryotic organization of their cells (Raven & Johnson, 1992). The contact of bacteria with other organisms is very common, and in plants and animals these associations may have cases of negative, neutral or positive consequences.

Plants are normally associated with different types of bacteria, externally (exophytic) and internally inside the body (endophytic). Plants or animals and bacteria may present a symbiotic relation that is the living together in close association of two dissimilar organisms in which both partners have advantages in living together. They can also present an endosymbiotic process when one of the partners live intimately and permanently inside the host (plant or animal) in a symbiotic association.

A great number of bacteria in nature have the ability to fix the nitrogen, which constitutes about 80% of the earth's atmosphere. This fixed nitrogen is very important for the survival of these bacteria and any excess can be and is frequently used by plants in different types of contact with these bacteria. For example, leguminous plants have a specific association with the nitrogen fixing bacteria of the genera Rhizobium as a symbiont, that is, it may live inside nodules in the root of leguminous plants but it can also survive and develop in nature in the soil. In each generation the nodules symbionts are produced by Rhizobium contamination through newly formed nodules in the plant roots.

Rhizobium in nature, when in contact with leguminous plants, is able to induce the formation of nodules in the root of the plant inside of which the bacteria develop, survive and produce nitrogen compounds which serve as fertilizer for the plant. This process has been known for more than a century and in the last decades it is in worldwide use in agricultural cultures with great success.

Nitrogen fixing bacteria endophytics, different from the Rhizobium group, living inside non-leguminous plants, were first described by Dobereiner and Pedrosa (1987) and presently there are several groups working in this field.

Until now, among the fixing nitrogen endophytic bacteria, there has not been found any association of a specific bacteria to a specific plant in such a way that the system could be exploited in agriculture. But even so this unspecified association is very important for the natural vegetation and for agriculture itself. As mentioned in the pre-introduction of this paper this is what we are working on in our laboratory in Brazil and we have found that a great number of the non-leguminous plants analyzed have endophytic nitrogen fixing bacteria associated with them.

The endosymbiotic bacteria in plants that we are analyzing in this article are found in the seeds and other parts of the organism and apparently they breed only inside the host and are transmitted to the next generation through the newly formed seeds.

We still are unable to determine whether in each seed there is only one or more than one type of this endosymbiotic bacteria. What we know is that every seed in dozens of plants which we have detailedly tested contains a great abundance of bacteria which are transmitted to some parts of the embryos in development, and on to the mature plant in which they are found in relatively small quantity in several parts of it and in the newly formed seeds in which they are more abundant.

In our experiment with bacteria living inside the plant seeds, frequently we may isolate more than one type although this does not mean that we may have more than one endosymbiotic bacteria in a simple seed. At the moment we still do not have enough data to distinguish the endosymbiotic from the simple endophytic and exophytic bacteria that also can exist in the seeds. The problem is that to isolate the bacteria from the seed we have to clean or decontaminate the seed with a solution of sodium hypoclorite (4%), alcohol (70%) and distilled sterile water.

Besides following the recommendations of the specialist we also doubled the time for this disinfecting but even so, we have found in several cases more than one species of bacteria isolated from a single decontaminated seed. We are now attempting to determine the taxonomic and molecular classification of these bacteria to be able to separate the endosymbiotic, endophytic and infectious bacteria coming from the seeds.

The association of a Leguminous plant and the nitrogen fixing bacteria Rhizobium is very important and well used in modern agriculture. The details of this association and of the nitrogen fixing process operated by the bacteria are today very well-known and used. It is of interest that together with this association with the Rhizobium the Leguminous plants also have an association with the Endosymbiotic bacteria, in this case not related to the fixation of nitrogen that we are describing in this article. Rhizobium are related to nodules produced in the root and only in the root of the plant, while the Endosymbiotic are more concentrated in parts of the seed and less frequent in other parts of the plant which is being investigated at the present time. In beans the contamination of the seeds come from bacteria present in certain parts of the flower. The flower produces the pod inside of which are the seeds. The bean seed is formed by two attached cotiledons having one embryo in between the two. In the embryos, formed by three sections attached together; one may distinguish two pre-leaves, one pre-stem and one pre-root. A great number of bacteria can normally be found in the cotiledons and embryos of the developed seeds. Details of their appearance and behavior may be seen in the www.eca.usp.br/nucleos/njr/pavan, which we are furnishing as part of this publication.

It is of interest to analyse the development of the seeds in peanuts (*Arachis hypogaea*), a plant of the Leguminosae family. We will transcribe a summary description of peanuts found in the Encyclopaedia Britannica: 'The flowers are borne in the axils of the leaves; what appears to be a flower stalk in a slender calyx up to 40 millimeters (1.6 inches) long. After pollination and the withering of the flower, an unusual stalklike structure called a peg is thrust from the base of the flower toward the soil. In the pointed tip of this slender, sturdy peg the fertilized ovules are carried downward until the tip is well below the soil surface. Only then does the tip start to develop into a characteristic pod where the seeds are formed.

Peanuts are a concentrated food; pound for pound, peanuts have more protein, minerals and vitamins than beef liver, more fat than heavy cream and more food energy (calories) than sugar'.

When referring to peanuts in the above comparison, in reality it is meant that the seeds of the peanuts are a concentrated food. Being so, it is of interest to know that these seeds have within their parts a great quantity of live bacteria which we are studying to find out if they are only a part of the food content or, what we think is more probable, if they are responsible for part of the process of food production or of the production of substances necessary for the development of the embryos.

Peanuts, being a Leguminous plant, are also a host to the bacteria Rhizobium in its root nodules that fix nitrogen through the synthesis of nitrogen compounds which are furnished to the plant as fertilizers.

In our experience with over one hundred seeds from different species and varieties of plants analysed, we may conclude that the presence of many bacteria inside the seeds is practically the rule.

In bird eggs, we analysed eggs of domestic chickens (about nine varieties), guinea hen, pheasant and quail. In all of them we found a similarity with our description for the domestic chicken and so we will discuss these most common chicken eggs for which endosymbiont bacteria are found in large quantities in the yolk. Since the chicken eggs have few cells, plenty of food reserved for the development of the embryos and a great number of bacteria, the egg phase is one in which the proportion of bacterial genetic material in relation to the chicken genetic material must be equivalent. This also is the situation in some plant seeds that we analysed.

Since we could not breed these bacteria in artificial medium we had to study the egg bacteria in their natural habitat. We began by boiling the eggs and observing what happens with the different components of the eggs and with the bacteria they contain.

One surprise was the ability of a ponderable quantity of bacteria to survive a submergence of the egg for 30 minutes in boiling water. Under these conditions the rest of the egg had the expected reaction: the yolk and the egg white coagulate normally. After separating the egg shell, the egg white and the yolk, a certain quantity of each isolated part of the cooked egg is mixed with distilled water and placed in a blender for a few minutes then subsequently placed in the autoclave (120°C for 20 minutes under pressure). These broths of the isolated parts of the chicken egg are individually mixed with a normal medium to make a combined medium in experiments to breed new bacteria.

Another surprise ensued when the autoclaved yolk broth was studied under a microscope. A substantial number of bacteria were still alive and moving normally in the space available on the microscope slide. It looks like these bacterias, although participating in their apparent activities, do not multiply regularly. This would be expected since the normal bacteria (the non autoclaved) in the artificial medium we offer them, do not divide themselves. It seems on the other hand that we have cases in which the autoclaved culture when left in glass tubes for a week or more shows an increased amount of the condensed material in the tubes. One possible explanation is that the bacteria lost the capacity of reproducing themselves but are still alive and active in the culture medium. We are testing these possibilities.

On the other hand, there are common cases in which certain bacteria are represented by colonies in which the individuals are attached to the neighbors forming a collar. These colonies in the microscope slide are in constant movement, whether in a concentrated colony or when the collar is stretched (see Home Page).

Collar-like colonies of bacteria can be found occasionally in non-autoclaved yolk but they are rare and the colonies may have a maximum of ten to fifteen individuals. Normally they move slowly and never as rapidly as the autoclaved ones.

We still do not know the origin of these bacteria and their roles in the autoclaved yolk medium.

The pieces of coagulated yolk that are present in the autoclaved yolk broth when examined under the microscope show beside its normal yolk compounds, aglutinated in pieces, some bubbles of gas and different pieces of gel or gum eliminated from the yolk, with many different and interesting figures. The gas and the gel figures are secreted by the coagulated yolk and at least part of these products appear to come from the coagulated yolk during the time that it is between the slides in the microscope.

We still cannot say if this elimination of gas and gel from the piece of autoclaved yolk constitutes material recently synthesized or had been there in a deposit inside the pieces of coagulated yolk. There are cases of movement and elimination of gas and gel 24 hours after the microscope slide was prepared. Very interesting are the great number of structures produced by the gum or gel moving free in the medium under microscope observation.

Of great interest, beside the shape and physical structure of the bodies observed in the microscope slide, is the movement that some of these bodies exhibit and for which at present we have no explanation (see Home Page).

In conclusion, we may say that, beside the scientific discovery of this new type of association between bacteria and plants and bacteria and birds, we expect that knowing the role of the bacteria in the seeds of plants and the eggs of birds will help to understand better the processes of development of the embryo and of the organism itself. Other important practical results can be obtained after elucidating the role these bacteria have in those processes.

Intracellular bacteria of the type we presented in seeds of plants and eggs of birds are very common in many species of Arthropods. The first case described by Hertig and Wolbach (1924) deals with a rickettsiae-like organism in reproductive tissues of the mosquito *Culex pipiens*. This bacteria was later described as *Wolbachia pipientis* by Hertig (1936) and after the eighties, other genera and species of intracellular bacteria were described and a great number of others Arthropods organisms and nemotods were shown to present the symbiotic association with this type of bacteria.

Different from what we know until now about the symbiotic association of bacteria with plants and birds, the bacteria of the Genus Wolbachia and the Cytophaga-like organism (CLO) associated with Arthropods induce several reproductive abnormalities like feminization, parthenogenesis, cytoplasmic incompatibility, these override the chromosomal sex determination, to convert infected genetic males into females and to induce embryonic lethality in the embryos that result when uninfected females are mated to infected males [Werren J.H. and Windsor D. (2000); Kostas, B. and Scoll, O'Neil (1998); Selivon, D. *et al.* (2002); Rousset, F. *et al*. (1992); Werren, J.H. *et al*. (1995); Weeks, A.R. *et al.* (in press); Kyei-Poku *et al*. (2005)].

It looks as though the symbiotic association of bacteria with plants and birds is more balanced than that which occurs in Arthropods. If anything similar to what occurs in Arthropods should have occurred in certain plants like beans, or in chickens it would have been discovered long ago. Genetic studies on normal beans Phaseolus or soya beans and domestic birds like chicken have been done for a long time and in a thorough way. If any of the endosymbiotic bacteria could cause in plants or birds one of the reproductive abnormalities that Wolbachia and related species may cause in Arthropods should have been discovered long ago. As a matter of fact I am surprised that these bacteria existing in such a large number and in so many plants and birds were not discovered until now.

It looks like the endosymbiosis which occurs in plants and birds are more balanced than the ones that occur in Arthropod. With the data we have until now, we will classify the endosymbiotic bacteria present in the seeds of plants and in the eggs of birds as friendly and auspicious partners.

REFERENCES

Hertig M. and Wolbach S.B. (1924), 'Studies on rickettsia-like microorganism in insects', *J. Med. Res*. 44, 329-74.

Hertig M. (1936), 'The rickettsia *Wolbachia pipientis* (gen. et sp. n.) and associated inclusions of the mosquito Culex pipiens', *Parasitology* 28, 453-86.

Kyei-Poku, G.K., Colwell, D.D., Coghlin, P., Benkel, B. and Floate, K.D. (2005), 'On the ubiquity and phylogeny of Wolbachia in lice', *Molecular Ecology* 14, 285-294.

Kostas, B. and O'Neil, S. (1998), 'Wolbachia Infections and Arthropod Reproduction', *Bioscience* 48 n. 4, 287-293.

Raven, P.H. and Johnson, G.B. (1992), 'Biology', *Publ. Morby Year Book*.

Rousset, F., Bouchon, D., Bernard, P., Juchault, P. And Solignae, M. (1992), 'Wolbachia endosymbionts responsible for various alterations of sexuality in arthropods', *Proc. R. Soc. Lond*. B. 250, 91-98.

Selivon, D., Perondini, A.L.P., Ribeiro, A.F., Marino, C.L., Lima, M.M.A. and Coscrato, V.E. (2002), *Invert. Reprod. and Develop*. 42, (2-3) 121-7

Weeks, A.R., Velten, R., Stouthamer, R., (in press), Dep. Entom. Univ. Calif. CA 92521, 0314 USA.

Werren, J.H. (1997), 'Biology of Wolbachia', *Annu., Rev. Entomol*. 42, 587-609.

Werren, J.H., Windsor, D. and Guo. L. (1995), 'Distribution of Wolbachia among neotropical arthropods', *Proc. R. Soc. Lond*. B 262, 197-204.

Werren, J.H. and Windsor, D. (2000), 'Wolbachia infection frequencies in insect: evidence of a global equilibrium', *Proc. R. Soc. Lond.* B 267, 1277-85.

Acknowledgments

My sincere thanks to Prof. Dr. Heloisa R. Barbosa, head of the Laboratory of Physiology of Microorganisms of the Biomedical Institute of the University of São Paulo, where in retirement I volunteered to study bacteria and I am enjoying my work with these friendly microorganisms. Many thanks also to Prof. Dr. Elizabeth Farrelly Pessoa for her help with the English version and to Paulo Pavan for his instructions in manipulating the computer to save the figures and upload them to our website.

MICROBIAL BIODIVERSITY: A NEW VOYAGE OF DISCOVERY

RAFAEL VICUÑA

INTRODUCTION

When I entered basic school exactly half a century ago, I was taught that all living organisms could be ascribed to either the animal or the plant kingdoms. Some years later, it had become evident that this was a rather simple way to assess the extraordinary richness of the biosphere. Hence, the former two kingdoms were extended to five: animals, plants, fungi, protists and monera (bacteria). Moreover, the members of the first four kingdoms could be grouped under the general category of the eukaryotes, with cells possessing membrane systems and organelles that are absent in bacteria. The latter are in turn grouped as prokaryotes. Of note is that this much broader classification does not include viruses. Whether these macromolecular structures should be considered living entities remains a matter of controversy until today.

Of these five kingdoms, plants and animals persist in concentrating most people's attention, including that of a large proportion of scientists. This attitude can be easily understood. When we contemplate nature, the immense variety of mammals, birds, fish, insects, trees, plants and flowers cannot be anything more than overwhelming. We spontaneously tend to associate biodiversity with the life forms that are accessible to our eyes. Biology textbooks contribute to this way of thinking, with illustrations of magnificent whales, lions, antelopes, zebras, tropical fish, ferns, old elms, butterflies, etc. The number of described species seems enormous to us: about one million arthropods, 270,000 vascular plants, 75,000 fungi, 70,000 mollusks, 45,000 chordates and 40,000 species of algae. In contrast, only 5,000 bacterial species have been identified.

However, the recent advent of powerful molecular techniques for the exploration of the microbial world is leading us to the realization that our criterion to assess biodiversity has been enormously restricted. First of all, we have learned that only between 0.1 and 10 per cent of bacteria normally present in the environment can actually be cultivated under laboratory conditions.[1] Therefore, any projection of total biodiversity on Earth that considered only those bacteria able to grow in culture had to be erroneous. Through both direct and indirect determinations, we are becoming aware that the dominance of bacteria in the biosphere is tremendous. In terms of biomass, the amount of carbon deposited in bacterial cells is at least equal, if not larger, to the total carbon present in animals and plants.[2] According to recent estimates, two-thirds of the bacterial biomass manage to survive within marine sediments, while the majority of the remainder reside in soil and the terrestrial subsurface. Most likely, large quantities of bacteria are also found in aquatic habitats, although they amount to only 1 per cent of the total.[3] To most people, it may come as a surprise to learn that an average human individual harbors as much as 1 kg of bacteria.[4] The number of bacterial cells contributing to this biomass is about 10^{14}, a figure roughly similar to that of the total number of cells composing the human body.

Paths of Discovery

Bacteria were first observed in 1676 by Antonie van Leeuwenhoek (1632-1723), a Dutch amateur microscope builder with little formal education. The primitive microscope designed by van Leeuwenhoek consisted of a lens mounted in a brass plate, adjacent to the tip of an adjustable focusing screw. The name for this novel instrument, although with a different shape, was proposed in 1625 by the physician-naturalist John Faber, who was a member of the *Accademia dei Lincei*: 'The optical tube ... it has pleased me to call, after the model of the telescope, a microscope, because it permits a view of minute things'.[5]

[1] Amann, R.I., Ludwig, W. and Schleifer, K-H., 'Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation', *Microbiol. Rev*. 59, 143-169 (1995).

[2] Whitman, W.B., Coleman, D.C. and Wiebe, W.J., 'Prokaryotes: The unseen microbial majority', *Proc. Natl. Acad. Sci. USA* 95, 6578-6583 (1998).

[3] *Idem*.

[4] Abbot, A., 'Gut Reaction', *Nature* 427, 284-286 (2004).

[5] Cited by Daniel J. Borstin in *The Discoverers* (Random House, NY, 1983).

The Royal Society of London, which by that time had been established in England for the communication of scientific work, published a series of letters describing van Leewenhoek's work, up until his death in 1723. Microscopic fungi had been discovered earlier by Robert Hooke (1635-1703), a British scholar of broad scientific interest and founding member of the Royal Society. Hooke, who also coined the word 'cell' for the first time after observing the texture of cork, published in 1665 a book entitled *Micrographia, or some physiological descriptions of minute bodies made by magnifying glasses with observations and inquiries thereupon.*[6] Apparently, this book had strong influence in van Leeuwenhoek's subsequent discoveries and therefore it would be fair to say that Hooke and van Leeuwenhoek share the credit for founding microbiology.[7]

Due to the lack of innovation in the design of new microscopes, early progress in microbiology was slow. It was only after the middle of the nineteenth century that a change in circumstances contributed to move forward this young science. The German botanist Ferdinand Cohn (1828-1898) made the first attempts to classify bacteria and introduced the use of cotton plugs to prevent contamination of sterile culture media. Also by that time, the French chemist Louis Pasteur (1822-1895) and the English physicist John Tyndall (1820-1893) provided definitive proof that micro-organisms do not arise as a result of spontaneous generation from lifeless matter. This was not the only controversial issue that Pasteur helped to clarify. He also demonstrated that decomposition of food and other organic materials were not merely chemical processes, but required the participation of micro-organisms. Moreover, fermentations revealed the remarkable fact that life can proliferate in the absence of oxygen. Tyndall and Cohn further discovered the formation of spores resistant to heat, a finding that led to the improvement of sterilization procedures.

However, proof that micro-organisms cause disease provided the greatest drive for the development of microbiology. Discoveries in sanitization procedures, made earlier by the British surgeon Joseph Lister, found their scientific support in 1876, through the conceptualization of the 'germ theory of disease' by Robert Koch (1843-1910). The so-called Koch postulates were formulated to demonstrate that a specific type of micro-organism causes a particular disease. Another key contribution of this German coun-

[6] J. Martín and J. Allestry, Printers to the Royal Society, London.
[7] Gest, H., 'The discovery of micro-organisms revisited', ASM News 70, 269-274 (2004).

try doctor was the development of inventive methods for obtaining the growth of bacteria in pure cultures. Previously, Pasteur had used rather simple liquid media to grow fermentative micro-organisms. However, Koch by adding new ingredients modified these media to establish nutrient broths, in both their liquid and solid versions, suitable for growing disease-producing bacteria.

While the role played by micro-organisms as agents of infectious diseases was of central interest in the late decades of the nineteenth century, studies on the participation of bacteria in the cycle of matter on earth began to receive growing attention. Relevant scientists in this issue were the Russian Sergius Winogradsky (1856-1953) and the Dutchman Martinus Willem Beijerinck (1851-1931), who discovered that the metabolic versatility of bacteria is much broader than that of plants and/or animals. Bacteria were shown to grow in completely inorganic cultures and to play an irreplaceable role in the carbon cycle. These scientists went on to show that bacteria are responsible for the fixation of atmospheric nitrogen into organic compounds. Another major contribution of Winogradsky and Beijerinck was the concept of enrichment culture, a miniature application of the principle of natural selection. It consists of a culture of defined composition that is inoculated with a sample containing a complex microbial population. After a few generations, those bacteria having the ability to grow faster in this specific medium will predominate. Enrichment cultures have been utilized since to isolate a large variety of bacteria with specific nutrient requirements.

The experimental methods developed by the afore-mentioned pioneers of microbiology allowed us to gain insight in a new world previously inaccessible to the naked eye. During the twentieth century we learned that bacteria show an amazing ability to adapt to all sorts of environmental conditions. They can grow at temperatures as high as 106°C (*Pyrolobus fumarii*) or as low as 4°C (*Polaromonas vacuolata*). Some require a very acidic pH to proliferate (*Picrophilus oshimae*), whereas others prefer extreme alkaline settings (*Natronobacterium gregory*). Bacteria have been shown to withstand pressures of up to 700 atmospheres (MT41) and to require salt concentrations above 25% (*Halobacterium salinarum*). Just a few months ago, *Nature* magazine reported the presence of a community of bacteria living in a lake beneath an Icelandic glacier.[8] This lake, placed inside the crater of

[8] *Nature* news published on line (13 July 2004).

the Grimsvötn volcano, is 100 meters deep and it is covered by 300 meters of a thick ice-sheet. This abundance of extremophiles makes us wonder about the possibilities of extraterrestrial microbial life. Paradoxically, there seems to be a place on the surface of the Earth where there is no life at all, namely, the Atacama Desert in Chile.[9]

Bacteria also became the organisms of choice to elucidate different metabolic pathways as well as to unveil the mechanisms of gene expression. These studies showed that in terms of metabolic versatility, bacteria are unsurpassed. As previously mentioned, only bacteria have the capacity to reduce the rather inert molecule of nitrogen, a task performed with a large expense of chemical energy. Most of the methane or 'natural gas' encountered in the outer few kilometers of the Earth's crust or in the atmosphere is the product of bacteria that use carbon dioxide as electron acceptor in their metabolism. On the other hand, bacteria conduct one-fifth of the photosynthesis occurring in the planet.[10] On a yearly basis, this represents twice as much as the fossil fuel energy used by human consumption the world over.[11] Some bacteria can obtain their energy by oxidizing inorganic molecules such as H_2, H_2S, Fe^{+2}, NH^{+4} and uranium^{+4}, among several others. Bacteria are able to consume the hydrocarbons present in oil, using them as carbon source for survival. They also have the capacity to degrade highly toxic xenobiotic compounds, such as pesticides, polychlorinated biphenyls, munitions, dyes and chlorinated solvents. Thus, as stated by the late Stephen J. Gould, 'bacteria represent the great success story of life's pathway. They occupy a wider domain of environments and span a broader range of biochemistries than any other group. They are adaptable, indestructible and astoundingly diverse'.[12]

A Hidden Microbial World

In spite of the great usefulness of pure cultures, the fact that most bacteria defy cultivation in the laboratory markedly restricted our knowledge of the microbial world. A key breakthrough which delivered a profound

[9] C. McKay, as quoted by J. Whitfield in *Nature* 430, 288-290 (2004).

[10] Dr. Dave J. Scanlan, personal communication.

[11] Lehiniger, A.L., Nelson, D.L. and Cox, M.M., *Principles of Biochemistry*, 2nd edition (Worth Publishers, New York, 1993).

[12] Gould, S.J., 'The evolution of life on earth', *Sci. Am.* 271(4), 85-91 (1994).

impact on our perception of microbial diversity took place about 30 years ago with the introduction of molecular sequences as a criterion to relate organisms. Because the number of sequence differences in a molecule is proportional to the number of stable mutational changes fixed in the encoding DNA, evolutionary distances can be measured by differences in the nucleotide or aminoacid composition in homologous nucleic acids or proteins, respectively. Based on this concept, the American microbiologist Carl Woese proposed the use of ribosomal RNA (rRNA) sequences as evolutionary chronometers. Ribosomes are cytoplasmatic particles made up of two subunits containing RNA and protein, which are part of the protein synthesizing machinery of the cell. Ribosomal RNAs constitute an adequate macromolecule for measuring phylogenetic distances because they are functionally constant, universally distributed and moderately well conserved. By aligning small-subunit RNA sequences from various organisms and counting the corresponding nucleotide differences among them, Woese constructed a phylogenetic tree that could be used to relate all organisms in the biosphere. Hence, it became known as the tree of life. Woese's tree showed three primary lines of evolutionary descent that span across the formerly known five kingdoms of living organisms. These lines are called urkingdoms or domains, and comprise the Eucarya (eukaryotes), Bacteria (previously called eubacteria) and Archea (initially called archeobacteria).

In the following years, complete genome sequences confirmed the existence of Archea as a separate evolutionary line that branched from the Eukarya. In addition, widespread sequencing revealed several short rRNA sequences that are unique to certain groups of bacteria. Linked to a fluorescent dye, these so-called signature sequences have been widely used as probes for the detection of bacteria (FISH: fluorescent in situ hybridization). Woese's tree has been now widely accepted, although it must be taken with some caution. First, due to multiple events of lateral gene transfer or even to intermixing of genomes in the course of evolution, the rRNA tree is not always congruent with phylogenetic trees based on metabolic genes. In addition, not all rRNAs are detected with the same sensitivity (the universal primers used in PCR experiments do not amplify all rRNA genes with the same efficiency). On the other hand, due to rRNA genes varying in number among prokaryotes, estimations of species diversity and abundance in a particular sample may not be highly reliable.

A simple glimpse at the tree of life makes it evident that the large organisms in which scientists have concentrated their attention constitute an extremely limited fraction of the diversity of life. For example, the relatedness of the human lineage to any kind of insect is closer than that of any two bacteria belonging to different phylogenetic groups. The tree also lends support to the hypothesis that mitochondria and chloroplasts derive from bacterial symbionts, most likely proteobacteria and cyanobacteria, respectively. On the other hand, the tree shows that the eukaryotic nuclear line extends as deep in the history of life as that of bacteria, and that it later gave rise to the archeal lineage.

By 1987, Woese and colleagues had completed a tree with cultivated micro-organisms delineating 11 major phyla or lineages,[13] which thereafter became 12 as a result of the separation of the gram-positive into two branches (*Firmicutes* and *Actinobacteria*). Among these original phyla are the gram-negative *Proteobacteria*, which include *Escherichia* and *Pseudomonas*, as well as the photosynthetic *Cyanobacteria*. Since then, about 14 additional phyla of bacteria growing in cultures have been defined. Several of these (*Thermodesulfurobacteria*, *Coprothermobacteria*, etc.) are thermophilic.

By the same time Carl Woese was sorting out his tree of life, his colleague Norman Pace proposed the use of modern molecular techniques to gain knowledge of the microbial diversity in natural environments. His rationale was based on the fact that with the sequence-based taxonomic criterion, a sole gene sequence suffices to ascribe an organism to a known phylum or to define a new one. Ribosomal RNA genes from uncultivated micro-organisms can be obtained directly from environmental samples by amplification with the polymerase chain reaction (PCR). Due to the relatively high conservation of rRNA genes, primers can be designed in such a way that they will anneal to sequences that are shared by representatives of all three domains. The fragments thus obtained are resolved by cloning and then sequenced. This approach has proven highly successful, leading to the identification of about 26 new phyla that contain no known cultivated representatives. These are found in a variety of habitats and some are highly abundant, especially in the previously unexplored Earth's crust. Most of the latter rely on energy provided by redox reactions of inorganic compounds, as opposed to those that depend either on the harvesting of sunlight or on the metabolism of organic compounds as sources of energy. Most often, the new lineages are distantly related to previously characterized ones and the shape of the tree reveals that bacterial diversity arose as a result of a radiation of lineages rather than from a sequential divergence from a main line.

[13] Woese, C.R., 'Bacterial evolution', *Microbiol. Rev*. 51, 221-271 (1987).

To-date, more than 60,000 small subunit rRNA sequences from a wide variety of prokaryotes have been reported.[14]

Further Innovations of the Molecular Methods of Analysis

The new molecular techniques proposed by Pace have become highly refined. They have even been focused to decipher functional information of the unmasked microbes, a characteristic that is rarely provided by phylogeny based on rRNA. In a recent publication,[15] Craig Venter and collaborators analyzed bacterial samples from the Sargasso Sea, a nutrient-limited, open ocean environment, by applying a whole genome shotgun sequencing method. In this approach, total DNA from a complex bacterial mixture is isolated and sheared into tiny fragments that are cloned and sequenced from both ends. Based on the sequences obtained, the fragments are pieced back together into their proper genomic arrangement with computer guidance. In this study, Venter and colleagues obtained more than 1,000 billion base pairs of non redundant sequences, equivalent to approximately 775 complete microbial genomes. Based on sequence relatedness, they estimated that the DNA analyzed was derived from 1,800 genomic species, including 148 previously unknown bacterial phylotypes. Other phylogenetic trees constructed with markers such as the elongation factor Tu or the heat shock protein 70 gave more conservative estimates, ranging between 341 and 569, of species richness.

This difference in the number of species obtained by using diverse criteria brings us to a complex issue: the definition of species in the prokaryotic world. For most animals, plants and fungi, a species corresponds to an interbreeding population. However, although bacteria interchange genetic material by various means, they do not have formal sex. Therefore, identification of bacterial species is a matter of consensus among microbiologists. A common rule of thumb is that two bacteria classify in the same species if their identity in rRNA is higher than 97%. Common phenotypic

[14] Cole, J.R., *et al.*, 'The ribosomal database project (RDP-II): previewing a new autoligner that allows regular updates and the new prokaryotic taxonomy', *Nucleic Acids Res*. 31, 442-443 (2003).

[15] Venter, J.C., *et al.*, 'Environmental genome shotgun sequencing of the Sargasso Sea', *Science* 304, 66-74 (2004).

characters and overall genomic coherence serve as additional criteria to confirm the rRNA sequence based diagnostic.

The study conducted by Venter and collaborators also led to the identification of 1.2 million previously unidentified predicted genes, of which about 70,000 are novel. This represents about an order of magnitude over the number of sequences presently archived in public databases such as SwissProt and REM-TrEMBL (about 140,000). The predicted genes could be classified by function, i.e. central intermediary metabolism, DNA metabolism, transcription, signal transduction, transport, etc. Due to the large heterogeneity of the microbial population, reconstruction of near complete genomes was not possible. However, the identity of some genes provided key information about the physiological characteristics of the community. For example, it identified the dominating cyanobacteria species performing photosynthesis (*Prochlorococcus*). Also, the identification of about 800 new genes encoding bacteriorhodopsin confirmed that coupling of light energy harvesting and carbon cycling through a non-chlorophyll based pathway is an important biochemical process in the ocean.

A similar shotgun approach can be applied for objectives that go beyond bacterial identification. For example, Gene Tyson *et al.*[16] have reassembled genomes from microbes occurring in an acidic ecosystem. Analysis of each reconstructed genome led to the identification of the corresponding bacteria, but also revealed pathways for carbon and nitrogen fixation, as well as for energy generation. It also provided insights into survival strategies in this extreme environment. In this particular case, this strategy was successful due to the small number of species population and the low frequency of genomic rearrangements, gene insertions and deletions. Sequencing of only 76 million base pairs sufficed to reconstruct two near complete genomes and three other partial genomes of the bacteria thriving in this habitat.

The culture-independent genomic analysis of microbial communities has been termed metagenomics.[17] It involves extracting DNA directly from some habitat, cloning it into suitable vectors and transforming it into a culturable host cell. If the aim is to study genome organization and genes encoding metabolic pathways, libraries are constructed in vectors that hold

[16] Tyson, G.W., *et al.*, 'Community structure and metabolism through reconstruction of microbial genomes from the environment', *Nature* 428, 37-43 (2004).

[17] Schloss, P.D., and Handelsman, J., 'Biotechnological prospects from metagenomics', *Curr. Opinion Biotechnol*. 14, 303-310 (2003).

large fragments of DNA that are subjected to sequencing. Alternatively, functional expression of individual genes can be attained by constructing libraries in expression vectors that allow insertion of the foreign gene next to a promoter. Thus, metagenomics aims to establish a link between phylogeny and function. Functional analysis has the potential to expand the range of known functions and elucidate functions of genes with no known homologs. Metagenomic analysis has been particularly useful to study soil micro-organisms. Several genes and gene products have been discovered in DNA extracted directly from soil, and of particular interest are those involved in pathways for biosynthesis of or resistance to several novel antibiotics.[18] In spite of the promising results obtained in this function-driven analysis, the construction of metagenomic libraries in expression vectors faces some drawbacks. Firstly, some *Bacteria* and *Archea* are resistant to lysis and therefore their DNA is not represented in the libraries. Secondly, some genes or their products are toxic to the host cell, typically *Escherichia coli*, or the foreign gene fails to be expressed due to its phylogenetic distance to the host cell machinery.

Concluding Remarks

New molecular techniques are manifesting the short-sightedness in our estimations of bacterial diversity and versatility. In spite of their small size and limited amount of genetic material, bacteria have evolved incredible sophisticated strategies to survive and proliferate in the most varying of environmental niches.

Analysis of DNA from uncultured micro-organisms has doubled the number of bacterial phyla and although major lineages will probably increase still further, the rate of expansion of the tree of life will probably be less explosive than in recent years. However, in spite of the importance of reaching an accurate knowledge of biodiversity, it is likely that new efforts will concentrate on the identification of novel genes. The ultimate goal should be to identify all genes in the biosphere. Because all bacteria share a certain percentage of genes, as more complete genomes and shotgun sequences become available, knowledge of the global genetic comple-

[18] Handelsman, J., *Soils – The metagenomic approach, Microbial biodiversity and bioprospecting*, A.T. Bull (ed.), ASM Press, Washington, pp. 109-119.

ment will be approached asymptotically. Major advances are expected to come in the functional predictions based on genome sequences. Although at first it may seem unobtainable to assign a specific role to each of such a vast array of novel genes, this task may be less daunting as this discovery of genes is likely to be far smaller than the variety of microbial species.[19]

We may wonder how realistic is it that we will ever reach an accurate assessment of Earth's microbial biodiversity. The massive lateral transfer of genes and the bacteria's rapid adaptation to an every changing environment represents great challenges for the future. However, it has become clear, on this new voyage to discovery, that we have encountered and will encounter still, more wonders than Antonie van Leewenhoek could ever have dreamed of, when he first took his primitive microscope and ventured upon this new world.

[19] Cases, I. and de Lorenzo, V., 'The grammar of (micro)biological diversity', *Environ. Microbiol.* 4, 623-627 (2003).

GEOMAGNETISM, ‘VACILLATION’, ATMOSPHERIC PREDICTABILITY AND ‘DETERMINISTIC CHAOS’

RAYMOND HIDE

1. Prologue

‘Discoveries’ – as the organisers of this symposium on ‘Paths of Discovery’ emphasise – ‘are at the basis of new knowledge’. Some discoveries are made upon verification or ‘falsification’ of a theory, but in many cases serendipity plays a key rôle. Then a discovery is made whilst something else is being sought but the scientific mind and intuition of the researcher become directed towards the unexpected.

Serendipity certainly featured in some of the main events outlined in this contribution to the symposium. They started in 1947 when P.M.S. Blackett, a cosmic ray physicist then at the University of Manchester, proposed a testable new theory of the Earth’s magnetism [1], which over the next few years he and colleagues succeeded in ‘falsifying’ by observation and experiment. The events ended in 1963 when E.N. Lorenz, a dynamical meteorologist at the Massachusetts Institute of Technology (MIT), published an account of his work on deterministic non-periodic fluid flow, motivated by his interest in the predictability of weather patterns. Lorenz’s paper [2] was later recognised by scientists in other disciplines and by mathematicians as a seminal contribution to what subsequently became known as ‘chaos theory’. This now influences ideas and methodologies in many branches of science and technology.

Linking these studies were quantitative laboratory experiments in which I discovered in spinning fluids subject to steady (thermal) forcing several nonlinear régimes of flow of varying degrees of complexity in their spatial and temporal characteristics, including (multiply-)periodic (‘vacillation’) and highly aperiodic (‘chaotic’) régimes. Undertaken from 1950 to 1953 at the University of Cambridge (and later repeated and their findings confirmed by D. Fultz and his colleagues at the University of Chicago), the experiments were motivated in the first instance by my interest in geomagnetism and motions in the metallic core of the Earth. But they were to attract the attention of meteorologists engaged in research on large-scale atmospheric motions and influence Lorenz's mathematical work on atmospheric predictability and nonlinear dynamical systems.

The present article is based on notes prepared originally in response to interest expressed by mathematicians and others in the geophysical background to the experiments.

2. Geomagnetism and Motions in the Earth's Liquid Outer Core

Speculations as to the origin of the Earth's magnetism go back several centuries, but geophysicists now agree that the phenomenon must be due to ordinary electric currents flowing within the Earth's metallic core, where they experience least resistance. Chemical and thermoelectric effects are unlikely to be strong enough to account for the electromotive forces needed to maintain the currents against ohmic dissipation, but motional induction involving hydrodynamical flow in the liquid outer core cannot be ruled out on quantitative grounds. This is the main reason why theoretical geophysicists – now equipped with powerful super-computers – have since the mid-1940s been prepared to wrestle with the mathematical complexities of 'self-exciting dynamos' in electrically-conducting fluids.

Dynamos convert the kinetic energy associated with the motion of an electrical conductor through a magnetic field into the magnetic energy associated with the electric currents thus generated in the moving conductor by the process of motional induction. In self-exciting dynamos permanent magnets are not necessarily involved; all that is needed is the presence of a very weak background magnetic field when the dynamo is started up. The self-excitation principle was discovered in the 1860s by engineers concerned with the development of practical systems of public electricity supply [3], who experimented with devices in which the rotating armature was connected by sliding electrical contacts to a suitably-oriented stationary field coil.

Such devices are topologically more complex in their structure than a continuous body of electrically-conducting fluid, such as the Earth's liquid metallic outer core. So it is by no means obvious that self-exciting dynamo action is possible in fluid systems, but it turns out to be true. From the equations of electrodynamics, theoreticians seeking mathematical 'existence theorems' have been able to show that in an electrically-conducting fluid most flows of sufficient rapidity and complexity in form are able to produce and maintain a magnetic field against ohmic dissipation. The first existence theorems [4] were produced, independently, by G.E. Backus and A. Herzenberg in the same year, 1958. This was nearly four decades after J. Larmor, in a paper on solar magnetism, had made the original suggestion that self-exciting dynamo action might be possible in a moving fluid [5]. His important idea appeared less attractive in 1934, when T.G. Cowling [6] showed that motional induction was incapable of maintaining magnetic fields of the limited class that possess an axis of symmetry. According to an aggrieved Larmor [7], even the Council of the Royal Astronomical Society were prompted by Cowling's *non*-existence theorem to respond negatively to his idea. Larmor saw on quantitative grounds that some kind of dynamo mechanism was needed to explain solar magnetism.

It was against this background of uncertainty that Blackett [1] in 1947 offered geophysicists a new theory of the origin of the Earth's main magnetic field. Significantly, this was done several years before rock magnetism studies had produced convincing evidence of polarity reversals of the field. Blackett was then the Head of the University of Manchester's large and lively Department of Physics (and due to receive a Nobel prize in the following year for his work on cosmic rays). According to his theory, which invoked an earlier suggestion associated with the names of H.A. Wilson and E. Schrödinger, [1, 8] the main magnetic fields of the Earth, Sun and any other rotating astronomical body were all manifestations of a new law of Nature, whereby any massive rotating body would be magnetic in virtue of its rotation. Its magnetic moment would be proportional to its spin angular momentum with the constant of proportionality equal to the square root of the universal gravitational constant divided by twice the speed of light, implying that if correct the theory would provide a basis for unifying the laws of gravity and electromagnetism.

E.C. Bullard quickly pointed out that the new theory could be tested by determining the vertical variation of the geomagnetic field in the upper reaches of the Earth. Blackett responded by setting up a research team under one of his staff members, S.K. Runcorn, charged with the task of measuring the field in deep coal mines. The theory was soon 'falsified' by the findings of the 'mine experiment' [9] and also by a direct laboratory experiment carried out by Blackett himself [10].

Geophysicists concerned with the origin of the main geomagnetic field were thus left with little choice but to confront the mathematical complexities of 'geodynamo' theory. These stem from the essential nonlinearity of the equations of magnetohydrodynamics (MHD) that govern flows in electrically-conducting fluids, a subject then in its infancy associated with the name of H. Alfvén. MHD phenomena such as self-exciting fluid dynamos abound in large-scale systems such as stars and planets, where typical values of the 'magnetic Reynolds number' $R=UL\mu\sigma$ can be high. (Here U is a characteristic flow speed, L a characteristic length scale, μ the magnetic permeability of the fluid and σ its electrical conductivity). But the scope for investigating such phenomena on the small scale of the terrestrial laboratory is very limited, owing to the difficulty with available conducting fluids of attaining high values of R.

Buoyancy forces due to the action of gravity on density inhomogeneities associated with differential heating and cooling produce fluid motions in stars and planets. The motions transfer heat by (free) thermal convection and their patterns are influenced by gyroscopic (Coriolis) forces due to general rotation. W.M. Elsasser pointed out in 1939 that the influence of Coriolis forces on convective motions in the Earth's liquid outer core may somehow account for the approximate alignment of the geomagnetic field with the Earth's rotation axis [11] – which for nearly a thousand years has been exploited by navigators using the magnetic compass.

3. 'Vacillation' and Other Régimes of Thermal Convection in a Rotating Liquid 'Annulus'

In 1948, as an impecunious undergraduate studying physics at the University of Manchester needing part-time paid employment, I joined the 'mine experiment' team as an assistant. The experience of working with Runcorn and his team stimulated my interest in geomagnetism and introduced me to the literature of the subject. Encouraged by Blackett and Runcorn, on graduating in 1950 I enrolled as a PhD student in the small Department of Geodesy and Geophysics at the University of Cambridge, where research in geodesy and seismology was already well established and new (and highly fruitful) initiatives were being taken in other areas – in marine geophysics by M.N. Hill and in palaeomagnetism by J. Hospers and Runcorn (who had moved from Manchester to Cambridge).

With some experience in experimental physics (but none in fluid dynamics), on reaching Cambridge I started some laboratory experiments on thermal convection in a cylindrical annulus of liquid (water) spinning about a vertical axis and subjected to an impressed axisymmetric horizontal temperature gradient. The necessary apparatus was quickly designed and constructed using equipment and other resources available in the department, including a war-surplus synchronous electric motor, a steel turntable used previously for grinding rocks, a supply of brass and glass tubing up to about 10 cm. in diameter and a recording camera incorporating a set of galvanometers which was no longer needed for field work in seismology. The resources also included, crucially, the facilities of a small workshop where research students could design and construct apparatus under the guidance of an experienced technician, L. Flavell.

My initial motivation amounted to nothing more than the hope that laboratory work on buoyancy-driven flows influenced by Coriolis forces due to general rotation might somehow shed light on motions in the Earth's liquid outer core. Luckily, promising lines of investigation emerged as soon as the apparatus was run for the first time, when a persistent regular flow pattern of four waves marked out by a meandering jet stream was seen at the top surface of the convecting liquid. By increasing the value of the steady angular speed of rotation of the apparatus, Ω (say), it was possible to increase the number of regular waves, M, but not beyond a point at which the pattern became highly irregular ('chaotic'). M could be decreased by reducing Ω, but not beyond a point at which the non-axisymmetric (N-) flow gave way to axisymmetric (A-) flow (see Figure 1 below).

The next steps were to investigate systematically how this behaviour depended not only on Ω but also on other impressed experimental conditions, namely the fractional density contrast ($\Delta\rho/\rho$) associated with the temperature difference maintained between the cylindrical side-walls of the 'annular' convection chamber, the depth (d) of the liquid within the 'annulus', and the width (b-a) of the gap between the side-walls – keeping the radius of curvature, b, of the outer side-wall fixed in the first instance. Empirical criteria were thus deduced for the occurrence of transitions (a) between the A-régime and the regular non-axisymmetric (RN-) régime, and (b) between the RN-régime and the irregular non-axisymmetric (IN-) régime.

The first of these transitions, (a), was found to occur at a critical value of the dimensionless parameter

$$\Theta = [gd\Delta\rho/\rho]/[\Omega^2(b\text{-}a)^2], \qquad (1)$$

where g denotes the acceleration due to gravity, which was typically much stronger than centripetal acceleration. The criterion indicates that loss of stability of the A-régime involves the conversion of potential energy into kinetic energy.

As to the criterion for the transition between the RN-régime and the IN-régime, the fully-developed regular waves of the RN-régime were found to be characterised by azimuthal wavelengths never exceeding approximately $3(b\text{-}a)/2$, with little dependence on d. The criterion implies a simple dependence of the value of M at the transition on the ratio $\Gamma=[b\text{-}a]/[(b+a)/2]$ and it indicates that the chaotic IN-régime ('geostrophic turbulence') arises when the RN-régime ('vacillation', see below) loses its stability through the non-linear transfer of kinetic energy between Fourier modes.

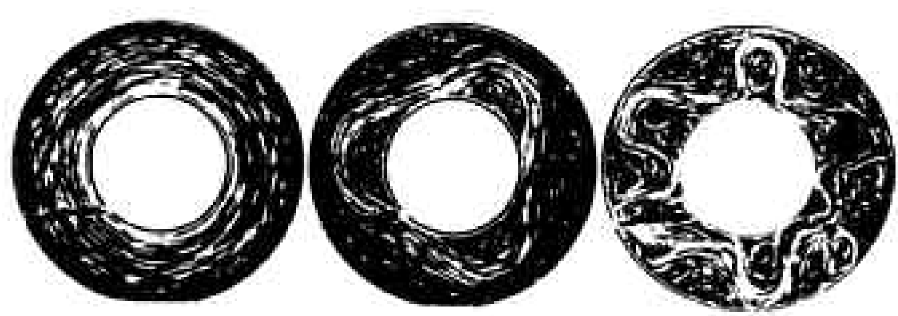


Figure 1. Streak photographs taken to illustrate three typical top-surface flow patterns, the first in the axisymmetic régime, the second in the regular non-axisymmetric régime (of 'vacillation') with M=3, and the third in the irregular ('chaotic') non-axisymmetric régime (of 'geostrophic turbulence'). The respective values of Ω were 0.34, 1.19 and 5.02 radians per second; other impressed conditions were held fixed.

Later experiments using glycerol/water mixtures indicated how these empirical criteria depend on the viscosity and thermal conductivity of the working liquid. The dependence is weak when Ω is so high that viscous effects are weak, but at low values of Ω the criteria exhibit dependence on the coefficient of viscosity, for which there is a critical value – found to depend on d, $(b\text{-}a)$ and Ω – below which axisymmetric flow occurs for all values of Θ.

The procedure followed in most investigations of the RN-régime involved setting Ω and other quantities required to specify the impressed experimental conditions at pre-determined values, and then waiting until transients had died away before measuring various properties of the flow that persisted. In some cases the persistent pattern of waves turned out to be steady (apart from a steady drift of the pattern relative to the annular convection chamber), but in others the pattern would undergo regular periodic fluctuations of various kinds. In the simplest of these the pattern exhibited pulsations in amplitude, which at their most pronounced were accompanied by alternations in the number of waves, M, from one cycle to the next. In other time-varying cases a sizeable local distortion of the wave pattern, sometimes amounting to the splitting of a wave, was seen to progress around the pattern, or the shape of the whole pattern would waver.

Significantly – in a manner reminiscent of the behaviour of a 'pin-ball' machine – when a number of experiments were carried out under the same impressed conditions there was a spread in values of M of the patterns that persisted, rather than a unique value of M. Thus, in a large number of trials under the conditions, say, of the second picture in Figure 1 (where M happens to be equal to 3), with each trial starting with the thorough stirring of the working liquid and then waiting for the resulting small-scale motions to die away, the resulting value of M of the persistent pattern that eventually formed would be equal to 2, 3 or 4, with relative probabilities depending on the value of the dimensionless parameter Θ.

Never expecting the term to stray beyond my laboratory notebook, I used 'vacillation' to denote the most extreme form of periodic 'wavering' seen in the experiments on the RN-régime. This occurred near the transition to the IN-régime. At one phase of the cycle the meandering jet stream gave way to separate eddies, which in turn decayed allowing the jet stream to reform, and so on. But when other workers took up annulus experiments (see below) some used the term 'vacillation' to signify *any* flow exhibiting regular periodic fluctuations. This made it necessary to introduce the terms 'shape vacillation', 'amplitude vacillation', 'wave-number vacillation', etc., leaving 'vacillation' on its own as an alternative term for the regular non-axisymmetric (RN-) regime (with *steady* non-axisymmetric flows as extreme cases when fluctuations are imperceptible).

Before leaving Cambridge in 1953 I completed my experimental work there by making – over substantial ranges of impressed conditions – further determinations of total convective heat transfer, flow velocities and patterns of temperature variations (using small arrays of thermocouples), and also of the non-unique dependence of M on Θ, etc., in the RN-régime. My main findings were summarised in two short papers [12], but several years (including a period of compulsory National Service) elapsed before any of the details of methods and results given in my PhD dissertation were submitted for publication in the open literature [13].

4. Geophysical and Astrophysical Fluid Dynamics and Dynamical Meteorology

General considerations of the dynamics of convective heat transfer in spinning fluids indicate that Coriolis forces promote departures from axial symmetry in systems characterised by axial symmetry in their boundary conditions [13]. The flow régimes found in the annulus exemplify this generic result, which has wide implications in 'geophysical and astrophysical fluid dynamics' (GAFD). And in view of the effective need for departures from axial symmetry that is implied by the existence theorems for self-exciting dynamos [5] and by Cowling's non-existence theorem [6], the result indicates one possibly key rôle played by Coriolis forces in the geodynamo process and the MHD of the Earth's core.

We note here, in passing, another phenomenon with wide implications in GAFD. This was observed during a brief study made of thermal convection in a rotating *spherical* (rather than cylindrical) annulus subjected to a horizontal temperature gradient and outlined in my PhD dissertation [13]. The study was intended to shed light on the effects on the pattern of motions in the Earth's liquid outer core that the presence of the underlying solid inner core might produce, thereby influencing details of the observed geomagnetic field. The experiments confirmed what general theoretical arguments predicted, namely that owing to Coriolis forces the most striking feature of the flow would be an extensive cylindrical 'detached shear layer' aligned parallel to the rotation axis and girdling the inner spherical surface, touching it along the equator. At sufficiently high rotation rates non-axisymmetric waves appeared on the detached shear layer.

But of more immediate significance during the course of the main experiments was the new dimension they acquired when the geophysicist and mathematician H. Jeffreys commented casually that some of my flow patterns reminded him of large-scale motions in the Earth's atmosphere. (Before losing interest in dynamical meteorology nearly two decades earlier, Jeffreys had made original contributions to the subject, starting when he was sent to work at the UK Meteorological Office during the First World War). So I started reading meteorological literature, handicapped at first by my inability to find dynamical meteorologists in Cambridge from whom I could obtain advice. The applied mathematicians there included several leading theoretical fluid dynamicists, but they found my experimental results 'mysterious'. They evidently preferred laboratory studies focused on the validation of mathematical analyses, at a time when many of the ideas and mathematical techniques needed for interpreting the essentially nonlinear behaviour exemplified by my results had yet to be developed.

However, I enjoyed helpful discussions about the atmosphere with E.T. Eady and other dynamicists in the Department of Meteorology at Imperial College, London. And in so far as subsequent developments along our 'path of discovery' are concerned, it was fortunate that around that time the director of the so-called 'Hydro Lab' of the Department of Meteorology of the University of Chicago, D. Fultz, was on sabbatical leave visiting fluid dynamicists and meteorologists in Europe. The Hydro Lab had been established a few years earlier at the initiative of two leading dynamical meteorologists – C.-G. Rossby of the University of Chicago and V.P. Starr of MIT – for the purpose of designing laboratory experiments that might shed light on the general circulation of the Earth's atmosphere.

Fultz told me about his careful literature search for relevant studies, in which he had uncovered reports of qualitative laboratory observations of flows in spinning fluids made by meteorologists F. Vettin (in 1857 in Berlin) and F.M. Exner (in 1923 in Vienna), whose findings had been confirmed by Fultz and his colleagues at the Chicago Hydro Lab in their so-called 'dishpan' experiments [14] – in which the convection chamber was an ordinary domestic aluminium open (American) dishpan. He was understandably interested in my work on the flow régimes obtained in the controllable, geometrically-simple and well-defined annulus apparatus, especially the regular non-axisymmetric régime. With a view to having my apparatus reproduced and my experiments repeated in his own laboratory, he visited me in Cambridge on several occasions in order to obtain details of my results and experimental techniques and of the design and construction of the rotating annulus.

Over the next few years (after I had left Cambridge and was engaged elsewhere in other work), the Hydro Lab confirmed my results and extended the experiments to somewhat lower rotation speeds than those used in the Cambridge studies [15]. And in his successful efforts to bring the experiments to the attention of other meteorologists, Fultz promoted the use of the term 'vacillation' and introduced nomenclature of his own. Thus, the critical dimensionless parameter Θ (see equation (1)) that I had deduced from my experimental data to be the main determinant of the characteristics of the annulus flows [12, 13] he termed the 'thermal Rossby number'; to my regular and irregular non-axisymmetric régimes of sloping convection he gave the single term 'Rossby régime'; and the axisymmetric régime he termed the 'Hadley régime' – after G. Hadley whose celebrated paper on the cause of the Trade Winds was published as early as 1735 [16].

Opinions still vary concerning the meteorological relevance of the laboratory experiments, but from an early stage Lorenz at MIT was amongst those who saw the importance of attempting to identify the dynamical processes underlying the various flow régimes, especially vacillation,[17] and exploring their implications for the predictability of atmospheric motions. To paraphrase his views as expressed in a monograph on the general circulation of the atmosphere [16]:

> So far as their meteorological significance is concerned the experiments, by indicating the flow patterns that can occur and the conditions favourable to each, have made possible the separation of essential from minor and irrelevant considerations in the theory of the global atmospheric circulation. They show, for instance, that while considerations of water vapour may yet play an essential rôle in the Tropics, it appears to be no more than a modifying influence in temperate latitudes, because the hydrodynamical phenomena found in the atmosphere, including even cyclones, jet streams and fronts, also occur in the laboratory apparatus where there is no analogue of the condensation process. The same remarks apply to topographic features, which were intentionally omitted in the experiments. The so-called 'beta-effect' associated with the sphericity of the spinning Earth – which produces a tendency for the relative vorticity to decrease in northward flow and increase in southward flow because of the variation with latitude of the Coriolis parameter – now appears to play a lesser rôle than had once been assumed. Certainly a numerical weather forecast would fail if the beta-effect were disregarded, but the beta-effect does not seem to be required for the production of typical atmospheric systems. The experiments have emphasised the necessity for truly quantitative considerations of planetary atmospheres. These considerations must, at the very least, be sufficient to place the Earth's atmosphere in one of the free non-axisymmetric régimes of thermal convection discovered in the laboratory work.

5. Theoretical Fluid Dynamics and Atmospheric Predictability

Theoretical work in fluid dynamics is based on the nonlinear four-dimensional (space and time) partial differential equations (PDEs) in terms of which the laws of dynamics and thermodynamics can be expressed mathematically. The equations of electrodynamics are also needed in cases of MHD flows in electrically-conducting fluids. Being highly intractable, the equations yield to traditional analytical methods only in simple special cases when nonlinear terms can be neglected or treated as small perturbations.

Recent years have witnessed impressive progress in the application of numerical methods of solution that exploit the power of modern supercomputers, with dynamical meteorologists in centres for weather and climate forecasting amongst those at the forefront of these developments.[18] But much more remains to be done before entirely trustworthy results become obtainable in this way.

The idea of calculating how the weather will evolve, by solving the equations of hydrodynamics using the meteorological data describing the present weather as the initial conditions, goes back to the work by V. Bjerknes and L.F. Richardson in the early twentieth century. A note of caution was issued at the time by H. Poincaré (whose mathematical work on the 'three-body problem' in planetary dynamics had introduced ideas and methods which are now used widely in chaos theory) when he wrote [19]:

> Why have meteorologists such difficulty in predicting the weather with any certainty? Why is it that showers and even storms seem to come by chance, so that many people think it quite natural to pray for rain or fine weather, though they would consider it ridiculous to ask for an eclipse (of the Sun or Moon) by prayer. We see that great disturbances are generally produced in regions where the atmosphere is in unstable equilibrium. The meteorologists see very well that the equilibrium is unstable, that a cyclone will be formed somewhere, but exactly where they are not in a position to say; a tenth of a degree (in temperature) more or less at a given point, and the cyclone will burst here and not there, and extend its ravages over districts it would otherwise have spared. If they had been aware of this tenth of a degree, they could have known of it beforehand, but observations were neither sufficiently comprehensive nor sufficiently precise, and that is why it all seems due to the intervention of chance.

When studying particular aspects of the behaviour of a fluid dynamical system, the governing nonlinear PDEs can be rendered less intractable, albeit less reliable, by simplifying the spatial and/or temporal representation of processes of secondary interest, as in the so-called 'intermediate' theoretical models. And in extreme cases such as the 'low-dimensional' theoretical models (sometimes called 'toy' models) employed when interest focuses on the influence of nonlinearity on temporal behaviour, further simplifications are effected when formulating the model by 'parameterising' all spatial structure. The resulting system is governed by ordinary differential equations (ODEs) needing comparatively modest computers for their analysis, but their solutions can have nothing more than a qualitative bearing on the prototype.

6. Low-Dimensional Models, the Lorenz Equations and Deterministic Chaos

Low-dimensional models bearing on the nonlinear behaviour of self-exciting fluid dynamos are provided by systems of Faraday-disk dynamos. The simplest versions are those introduced in the 1950s by Bullard and T. Rikitake [20]. The autonomous set of nonlinear ODEs in three time-dependent variables that govern the Rikitake system of two coupled disk dynamos was shown in 1962 by D.W. Allan to possess persistent non-periodic (i.e. chaotic) solutions [21]. However, the character of these persistent solutions depends critically on the neglect of mechanical friction in the original Rikitake (and Bullard) systems.

In concurrent research, Lorenz was developing ideas about the use of low-dimensional models in the interpretation of vacillation[17] and other laboratory flow régimes and also about effects of nonlinear processes on atmospheric predictability [2]. His studies of the nonlinear amplification of the effects of tiny errors in meteorological data and its likely consequences for weather forecasting gave rise to the now-familiar term 'butterfly effect', which attracted the attention of writers on popular science as wide interest later developed in the subject of chaos. Using mathematical and computational techniques he investigated a low-dimensional 'toy' model of convection governed by what later became known as the 'Lorenz set' of three (dimensionless) autonomous ODEs, namely:

$$\mathrm{d}x/\mathrm{d}t=a(y-x), \quad \mathrm{d}y/\mathrm{d}t=bx-y-xz, \quad \mathrm{d}z/\mathrm{d}t=xy-cz, \tag{2}$$

which contain two simple nonlinear terms, $-xz$ and $+xy$. Here $x(t)$, $y(t)$ and $z(t)$ are the three time (t)-dependent variables and a, b, and c are positive 'control parameters'. In one of his solution régimes Lorenz found non-periodic behaviour that would be termed 'deterministic chaos' nearly a decade later in mathematical work on nonlinear dynamical systems. Through its impact on the development of ideas in the theory of such systems, the published account of Lorenz's work [2] became one of the most influential scientific papers of the past few decades [22].

The Lorenz equations and other sets of autonomous nonlinear ODEs continue to provide fruitful lines of mathematical research [22]. And in the words of J.D. Barrow [23] writing about the influence of chaos theory on mathematics:

> The mainstream of mathematics has begun to move away from the high ground of extreme formalism to the study of particular problems, notably those involving chaotic nonlinear phenomena, and to seek motivation from the natural world. This is a return to a distinguished tradition for ... there are complementary examples where our study of the physical world has motivated the invention of new mathematics. The contemplation of continuous motion by Newton and Leibniz ... led to the creation of the calculus ... (and) Fourier series arose from the study of heat flow and optics. In the twentieth century, the consideration of impulsive forces led to the invention of 'generalised functions' ... (which) were used most powerfully by Paul Dirac in his formulation of quantum mechanics. ... In recent years this trend towards specific applications has been perpetuated by the creation of a large body of dynamical systems theory, and most notably the concept of a 'strange attractor', as a result of a quest to describe turbulent fluid motions. The growing interest in the description of chaotic change, which is characterised by the very rapid escalation of any error in its exact description as time passes, has led to a completely new philosophy with regard to the mathematical description of phenomena. Instead of seeking more and more mathematical equations to describe a given phenomenon, one searches for those properties which are possessed by almost every possible equation governing change. Such 'generic' properties, as they are called, can therefore be relied upon to manifest themselves in phenomena that do not possess very special properties. It is this class of probable phenomena that are most likely to be found in practice.

7. Nonlinear Stability and Quenching

The disorder and associated lack of predictability of motions in the Earth's atmosphere and also of flows encountered in other nonlinear fluid systems – such as Lorenz's toy model in the chaotic régime [2] and the laboratory annulus in the irregular non-axisymmetric régime [12, 13] – are due to instabilities associated with feedback and coupling. But nonlinear processes can in some circumstances promote *stability* and *order*, rather than instability and disorder.

Such behaviour can be investigated by modifying the feedback and coupling terms in well-known autonomous sets of nonlinear ODEs [24]. Denote by V the 'volume' of that region of $(a,b,c,$etc.$)$ 'parameter space' where instability of equilibrium solutions gives rise to persistent solutions that fluctuate either periodically or non-periodically (i.e. chaotically) and consider the sets obtained by multiplying each of the nonlinear terms in equations[2] by a 'quenching function' q (say). In general $q=q(x,y,z)$, with $q=1$ corresponding to the special case of the Lorenz set. In the representative cases when $q=1\text{-}e+ey$ with e ranging from 0 to 1, V decreases monotonically with increasing e and vanishes when $e=1$. Fluctuating persistent solutions are then completely quenched for all values of (a,b,c), leaving only stable steady equilibrium solutions throughout the whole of (positive) (a,b,c) parameter space [24]!

Nonlinear quenching of the chaotic behaviour of the geodynamo associated with modest changes in boundary conditions at the surface of the Earth's liquid core has been invoked to account for the intermittency seen in the irregular time series of geomagnetic polarity reversals over geological time, with intervals between reversals varying from 0.25MY to 50MY [23]. And there are other examples of nonlinear processes promoting stability rather than instability. Such processes underlie the stability of annulus flows in the régime of vacillation, the comparative regularity of large-scale motions in the atmosphere of the planet Mars and the durability of the Great Red Spot and other long-lived eddies in the atmosphere of Jupiter [12, 13, 25].

8. Epilogue

Research environments changed significantly over the four decades since the final stage of our chosen 'path of discovery' was reached, in 1963. Few areas of science have been left untouched by the astonishing growth in power and availability of computers, which now support most research projects including laboratory work on fluid flows and other nonlinear systems. Over the same period, new observations covering many wavelengths in the electromagnetic spectrum, made not only with ground-based instruments but also with instruments mounted on spacecraft, have had a major impact on meteorology, geomagnetism and other geophysical sciences.

Observations of the atmospheres of other planets (Venus, Mars, Jupiter, Saturn, Uranus and Neptune) now influence research in terrestrial meteorology and climatology, just as observations of the magnetic fields of other planets (Mercury, Jupiter, Saturn, Uranus and Neptune) – none of which had been discovered in 1947 at the start of our 'path of discovery' – influence research in geomagnetism. Larmor's prescient views on solar magnetism have been abundantly vindicated by subsequent research [26].

Our 'path' started with the publication of Blackett's theory of the Earth's magnetism, which was testable and timely. Even though the theory turned out to be wrong it led to important new work in other areas of geophysics. His 1947 paper [1] marks the starting point of yet another (better-known) 'path of discovery'. This involved investigations of the magnetism of rocks taken up in the early 1950s by two groups, one at Cambridge led by Runcorn and the other at Imperial College led by Blackett and J.A. Clegg. Using magnetometers of various types – including the highly sensitive astatic magnetometer designed initially by Blackett for testing his theory [10] – both groups investigated fossilised magnetic field directions of igneous and sedimentary rocks collected from several continents. This enterprise provided new evidence in support of ideas concerning continental drift put forward much earlier, in 1915, by A. Wegener, thereby advancing the general acceptance of the ideas by geologists [27] and setting the scene for the emergence towards the end of the 1960s of the remarkably successful theory of plate tectonics.

A brilliant and versatile physicist, Blackett encouraged basic and applied research in all branches of his subject. Many still remember a talk given in 1948 to a student society during which Blackett gave a clear and convincing explanation of the essential physics of magnetohydrodynamic waves, at a time when Alfvén's important new ideas – which in 1970 were recognised by the award of a Nobel Prize – had yet to gain wide acceptance. Those of us lucky enough to hear him lecture at early stages of our careers gained useful insights into the world of physics, and those who would later venture into research were also influenced by his remarks on areas worth avoiding.

The proceedings of this symposium on 'paths of discovery' are expected by the organisers to interest those concerned with the planning of programmes of research. In such exercises it is never easy, of course, to allow for serendipity, making the ideal of moving along 'well-illuminated open paths' rarely achievable in practice. But useful lessons will doubtless be learnt, even though progress towards a discovery often seems like 'moving around in a darkened room and bumping into furniture'.

REFERENCES

1. Blackett, P.M.S., 'The magnetic field of massive rotating bodies', *Nature 159*, 658-666 (1947); *Phil. Mag. 40*, 125-150 (1949).
2. Lorenz, E.N., 'Deterministic non-periodic flow', *J. Atmos. Sci. 20*, 130-141 (1963).
3. Bowers, B., *A history of electric light and power*, Stevenage, UK: Perigrinus Ltd. (1986); Jeffrey, J.V., 'The Varley family: engineers and artists', *Notes Rec. Roy. Soc. London, 51*, 263-279 (1997).
4. Backus, G.E., 'A class of self-sustaining dissipative spherical dynamos', *Ann. Phys. (NY) 4*, 372-447 (1958); Herzenberg, A., 'Geomagnetic dynamos', *Phil. Trans. Roy. Soc. London A250*, 543-585 (1958).
5. Larmor, J., 'How could a rotating body such as the Sun become a magnet?', *Rept. Brit. Assoc*., 159-160 (1919).
6. Cowling, T.G., 'The magnetic fields of sunspots', *Monthly Notices Roy. Astron. Soc. 94*, 39-48 (1934).
7. Larmor, J., 'The magnetic fields of sunspots', *Monthly Notices. Roy. Astron. Soc. 94*, 469-471 (1934).
8. Schröder, W. & Treder, H-J., 'Einstein and geophysics: valuable contributions warrant a second look', *Eos, Trans. Amer. Geophys. Union* 78, 479-485 (1997).
9. Runcorn, S.K., Benson, A.C., Moore, A.F. & Griffiths, D.H., 'Measurements of the variation with depth of the main geomagnetic field', *Phil. Trans. Roy. Soc. A244*, 113-151 (1951).
10. Blackett, P.M.S., 'A negative experiment relating to magnetism and the Earth's rotation', *Phil. Trans. Roy. Soc. A245*, 309-370 (1952).
11. Elsasser, W. M., 'On the origin of the Earth's magnetic field', *Phys. Rev. 55*, 489-498 (1939).
12. Hide, R., 'Some experiments on thermal convection in a rotating liquid', *Quart. J. Roy. Meteorol. Soc. 79*, 161 (1953); 'Fluid motions in the Earth's core and some experiments on thermal convection in a rotating fluid', 101-116 in *Proceedings of the First Symposium on the Use of Models in Geophysical Fluid Dynamics (Baltimore 1953)*, R.R. Long (ed.), (1953).
13. Hide, R., *Some experiments on thermal convection in a rotating liquid*, PhD dissertation, University of Cambridge (1953); 'An experimental study of thermal convection in a rotating liquid', *Phil. Trans. Roy. Soc. London A250*, 414-478 (1958).

14. Fultz, D., 'Experimental analogies to atmospheric motions', 1235-1248 in *Compendium of Meteorology*, American Meteorological Society (1951).
15. Fultz, D., Long, R.R., Owens, G.V., Bowen, W., Kaylor, R., & Weil, J., 'Studies of thermal convection in a cylinder with implications for large-scale atmospheric motions', *Meteorol. Monogr*. 4 (No. 21), 104 pages (1959).
16. Lorenz, E.N., *The nature and theory of the general circulation of the atmosphere*, Geneva: World Meteorological Organization (1967).
17. Lorenz, E.N., 'The mechanics of vacillation', *J. Atmos. Sci. 20*, 448-464 (1963).
18. Wiin-Nielsen, A.C., 'Models, predictions and simulations', 55-75 in *Changing concepts of Nature at the turn of the Millennium*, Hide, R., Mittelstrass, J. &. Singer, W.J. (eds.), Vatican City: Pontifical Academy of Sciences Press, Scripta Varia *95* (2000); Tong, H., (ed.), *Chaos and forecasting,* (Proceedings of a Royal Society Discussion Meeting, see especially article by Palmer, T.N., Buizza, R., Molteni, F., Chen, Y.Q. & Corti, S.), Singapore: World Scientific (1995); Norbury, J. & Roulstone, I. (eds.), *Large-scale atmosphere-ocean dynamics: Analytical methods* (vol. 1); *Geometric methods and models* (vol. 2), Cambridge University Press (2002).
19. Lighthill, M.J., 'The recently recognized failure of predictability in Newtonian dynamics', 35-50 in *Predictability in Science and Society*, Mason, J., Matthias, P. & Westcott, J.H. (eds.), Cambridge University Press (1986); Gribbin, J., *Deep simplicity; chaos, complexity and the emergence of life*, London: Allen Lane Penguin Books (2004).
20. Bullard, E.C., 'The stability of a homopolar dynamo', *Proc. Cambridge Phil. Soc*. 51, 744-760 (1955); Rikitake, T., 'Oscillations of a system of disk dynamos', *Proc. Cambridge Phil. Soc. 54*, 89-105 (1958).
21. Allan, D.W., 'On the behaviour of systems of coupled dynamos', *Proc. Cambridge Phil. Soc. 58*, 671-693 (1962); Ershov, S.V., Malinetskii, G. G. & Ruzmaikin, A.A., 'A generalized two-disk dynamo system', *Geophys. Astrophys. Fluid Dyn. 47*, 251-277 (1989); Hide, R. 'Structural instability of the Rikitake disk dynamo', *Geophys. Res. Letters 22*, 1057-1059 (1995).
22. Ruelle, D. & Takens, F., 'On the nature of turbulence', *Commun. Math. Phys. 20*, 167-192 (1971); Sparrow, C., *The Lorenz equations; bifurcations, chaos and strange attractors,* New York: Springer-Verlag (1982); Lorenz, E.N., *The essence of chaos*, University of Washington Press,

(1993); Thompson, J.M.T. & Stewart, H.B., *Nonlinear dynamics and chaos (second edition),* Chichester: John Wiley & Sons, Ltd. (2002).
23. Barrow, J.D., *Theories of everything; the quest for ultimate explanation*, Oxford University Press (1990).
24. Hide, R., McSharry, P.E., Finlay, C.C. & Peskett, G., 'Quenching Lorenzian chaos', *Int. J. Bifurcation and Chaos 14*, 2875-2884 (2004); Hide, R., 'Nonlinear quenching of current fluctuations in a self-exciting homopolar dynamo', *Nonlinear Processes in Geophysics 4*, 201-205 (1997); Hide, R., 'Generic nonlinear processes in self-exciting dynamos and the long-term behaviour of the main geomagnetic field, including polarity superchrons', *Phil. Trans. Roy. Soc. London A 358*, 943-955 (2000).
25. Hide, R. & Mason, P.J., 'Sloping convection in a rotating fluid', *Advances in Physics 24*, 47-100 (1975); Hide, R., Lewis, S.R. & Read, P.L., 'Sloping convection; a paradigm for large-scale waves and eddies in planetary atmospheres', *Chaos 4*, 135-162 (1994).
26. Weiss, N.O., 'Dynamos in planets, stars and galaxies', *Astronomy & Geophysics 43*, 3.9-3.14 (2002); Tobias, S. & Weiss, N.O., 'The puzzling structure of a sunspot', *Astronomy & Geophysics 45*, 4.28-4.33 (2004).
27. Blackett, P.M.S., Bullard, E.C. & Runcorn, S.K. (eds.), 'A symposium on continental drift', *Phil. Trans. Roy. Soc. London A258*, 1-323 (1965).

TODAY THE WORLD OF TOMORROW – THE ENERGY CHALLENGE

CARLO RUBBIA

First of all let me say that this is the last presentation and it is therefore slightly different from the previous one. I did not feel like coming down and telling you about my own path of discovery, maybe it is a form of modesty or whatever, however, it is certainly true that this Council has been devoted to such a topic. Now, there is no doubt that new knowledge is driven by the discovery action. The many presentations we have heard over the last few days have well illustrated the different ways in which scientific progress may manifest itself, either through individual researchers or, as it is done more often today, through research programmes involving many researchers often from very different disciplines. Indeed, discovery has progressively transformed itself from a separate action of a single individual to a collective result of a wider community.

In the most advanced part of mankind, research initially from the isolated endeavour of a few scholars, has now become the main engine of social and economic progress of the society as a whole. We are witnessing the progressive transformation towards what is called a knowledge-driven economy with the discovery process being the main engine of progress.

The social consequences of such a deep evolution should not be underestimated. However, not everybody will profit from this immense gift to mankind represented by the scientific and technological progress. What I could call the scientific and technological illiteracy is becoming a persistent problem for a vast component of the world population. It is one of the main responsibilities of the more advanced societies, the one of providing free access of such a knowledge-driven progress to the most needy ones, today so essential in order to heal most of the crucial problems of the poor, like illnesses, poverty, lack of food, lack of water and so on. Contributing to the solution of these injustices is today configured as one of the most important missions to which the scientific community must actively contribute. In addition to these traditional problems a new situation is progressively emerging, initially due to the rapidly expanding number of individuals on earth which are now 6 billion people and presumably twice as many in not too distant a future, namely a rapid growth of extremely serious climatic changes as described last Saturday by Professor Crutzen. Such a presumable change in earth climate ahead of us, of which we detect only the first signs without the most serious consequences for the poorest part of the planet, since they are the least prepared to cope with such major changes which are forecast ahead of us. And indeed we should realise that 95% of such climatic changes are due to energy production. Energy supply has been a major element in our civilisation. Historically, energy for food gathering has been supplemented by energy for household use, initially heating, to organise our culture, industry and transportation. The total energy consumption of the most advanced part of mankind has grown about 100-fold from the beginning of history, reaching today the level of about 0.9 gigajoules per day per person, about one gigajoule to each one of us every day. This corresponds to the equivalent of burning 32 kg of coal per day per person or a continuous average supply of 10 kilowatts of power per person. Hence, the basic food supply represents today, for most of us, only a mere 1% of the total energy needed by us.

A most remarkable feature of the pro-capite energy consumption is the disparity determined by the differences in social progress. The present enormous disparity in electric energy consumption – Sweden 15,000 kWh of electricity per person per year, Tanzania 100 kWh per person per year – demonstrates a huge correlation between energy and poverty. But there is no doubt that the world's energy will continue to grow in the future since the population is steadily increasing and billions of people in the developing countries are striving for a better life. Hopefully the disparity in energy consumption may tend to converge. According to IEA World Energy Outlook, about 1.6 billion people, a quarter of the current world's population, are without electricity, which precludes the great majority of industrial activities and the related job creation. The majority (4/5) of these populations live in rural areas in the developing countries, mainly in Asia and Africa. About 2.4 billion people rely almost exclusively on traditional biomass as their principal energy source. Incidentally, in many of those countries the level of solar flux is such that it could potentially become a new primary energy source provided it is harnessed with simple and cost-effective technology.

It is well-known every bit of this energy, if produced by burning fossils, is multiplied by the sun as much as over a hundred times because of the increased CO_2 radiative forcing caused by the persistent trapping of the sun's radiation by the burnt fossils in the earth's atmosphere. In other words, the energetic toll to the planet is about a magnitude greater than the initial heat generated by man. So we have 1 but we produce 100. In the past, before the beginning of the seventies, in a first approximation the carbon cycle was closed in an equilibrium situation until human activities started to tilt the balance. Presently, CO_2 emissions are about 6 gigaton of carbon equivalent, namely 22 gigaton of CO_2, growing at about 2% per year. During the last 10 years emissions were 63 gigaton carbon, corresponding to 32 accumulated in the atmospheres and the remaining 30.4 absorbed by the ocean and by vegetation.

In comparison, the simple breathing of 6 billion people alone produces already about 1 gigaton of CO_2 yearly. We can predict with confidence that in the next centuries the continuative use of fossils without restriction will dramatically modify the earth's climate in ways which will impact in practice every living organism. Technological improvement will no doubt introduce other, more acceptable, forms of energy but the planet is notwithstanding continuing to burn a significant fraction of fossils for a long time to come, especially in those parts of the planet where technological change is slowest. The estimated reserve of fossils are about 500 gigaton carbon for oil and 5,000 gigaton carbon for coal. Coal reserves could be larger by a factor of 2 or 3 if also less noble forms of energy would be burned. Likewise recovery and new discoveries may contribute with substantial increases in oil and natural gas. The fact of the cumulative emission of as many as 5,000 gigaton of carbon of natural coal, progressively burned by people, will depend on which rate it is burned with a maximum CO_2 concentration which will be about 4 times the present level, presumably reached somewhere between 400 and 800 years from today. Since the recovery time is very long and what matters is only accumulated concentration, the result is only slightly dependent on the actual distribution of emission. Let us assume for instance that the fossil consumption of as much as 5,000 gigaton carbon, although being concentrated in only some parts of the world, continues for something of the order of 800 to 1,000 years to come, as it takes to use all the available carbon. Two thousand years from today the CO_2 concentration is still likely to be twice the present level. I must say that half of the coal flames produced by the burning of Caligula's Rome fire are still present in the atmosphere and are still taking away from us CO_2.

At the present consumption level, known reserves for coal, oil, gas and nuclear correspond to a duration of the order of 230, 45, 63 and 54 years. This effect will be affected positively by discovering new reserves, negatively by increased consumption. Even if these factors are hard to assess, taking into account the long lead time for the development of new energy sources, the end of the fossil era is at sight. And what after that? It may be worth mentioning that only 2 natural resources have the capability of a long-run energetic survival of mankind, beyond fossils, which incidentally as mentioned may be prematurely curbed by intolerable climatic changes. They are 1) solar energy and 2) an innovative, different nuclear energy.

The present world's energy consumption is about 1/10,000 of the solar energy available on earth's surface. On 1 m^2 in a good location (sun belt), it 'rains' yearly the equivalent of = 25 cm of oil. Several forms of renewable energy may bring major progress, provided an adequate level of research and development is becoming available in the near future: biomass, wind energy and especially a highly innovative technology based on high-temperature solar heat which appears capable of a major breakthrough in the field of energy production. This technology, based on the use of a simple concentrated mirror, following an ancient method of Archimedes of Syracuse, may produce vast amounts of both electricity and synthetic hydrogen from water splitting with high efficiency. For instance, the surface theoretically required to generate the full planetary electricity demand in 2050 represents the solar energy extracted from an area of 200 times 200 km square, somewhere in the vast equatorial sunbelt region. In comparison, the area today dedicated to agriculture is two orders of a magnitude larger, namely ten to the seven km square. A second alternative is a new form of nuclear energy from natural uranium, thorium or lithium or some other light element which is called fusion. If burned, it may be adequate for many thousands of years of several times the present energy consumption. It is still a very long range of development, it is unlikely that the practical industrialisation of such a very sophisticated technology may occur within the time-span of even the youngest amongst the presently active individuals. However, the necessity for the future generation of a clean and inexhaustible source of energy justifies why so many people are spending so much effort towards such a goal. One should primarily target new alternative methods free from the present fundamental drawbacks like radioactivity in the fuel, radioactive waste and, most important, proliferation potentials which are the main environmental, political and social problems facing today's nuclear power, in this way clearing the way for the widespread nuclear plants especially in the developing world.

To conclude, a coherent energy policy and strategic choices have to be made on nuclear power relying primarily on innovative scientific and technological discoveries in order to reconcile sustainable development and economic growth with a threat of environmental decay. Energy supply has been a major element in our civilisation. No doubt, the long-range future of mankind would be impossible without a continued supply of plenty of energy. The problem of energy focuses today the interests of politicians, businessmen, technologists and people at large. Everybody will agree on the fact that energy is an absolute necessity for the future of mankind, but the consequences of an ever-expanding energy demand should not be underestimated since they represent a growing concern for the long-term future of mankind, both in terms of environmental problems and of the availability of supply. Our society will depend crucially on an uninterrupted and differential supply of plenty of energy, therefore major steps have to be taken in order to avoid potential geopolitical and price vulnerability conflicts.

The inevitable growth of energy consumption under the sheer momentum of society and the very human expectations of the poor may indeed add enough yeast to make this aspect leaven beyond control. I believe however that like in the case of famine, illness, etc., also here science and technology should be trusted. Indeed, there are reasonable expectations that combined they will have the possibility of solving also this problem in full accord with economic, dynamic and technical constraints that the working system has to comply with.

TABLES

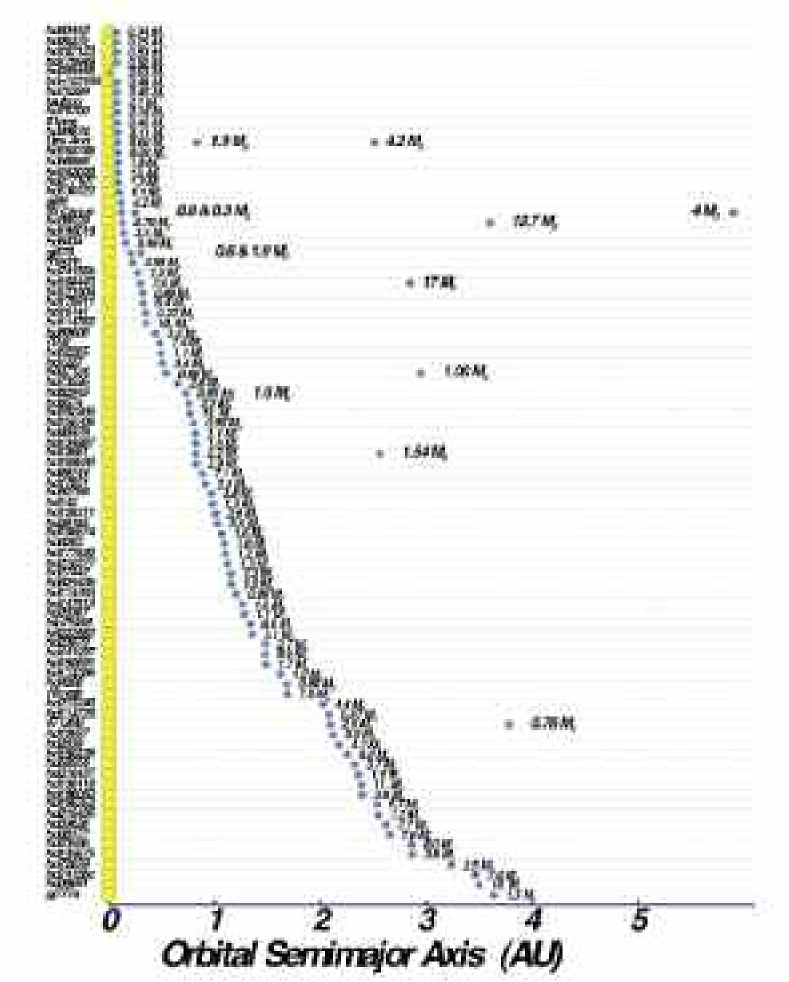


Figure 1. An overall view of the exoplanets known at the date September 2004 (Marcy *et al.*, Ref.[3]). Star names are on the left, distance to the star (semi-major axis of elliptical orbit) is in abcissa, measured in astronomical units (1 a.u. = 150 millions kilometers = distance Sun-Earth), planets are labeled with their mass, measured in units of the Jupiter mass M_J ($1\ M_J = 1/1000\ M_{\text{Sun}} = 326\ M_{\text{Earth}}$). Except in a few cases (transiting planets), these masses are lower limits, because of the uncertainty, due to the spectrometric detection method, on the inclination of the planetary orbit w.r.t. the plane of the sky.

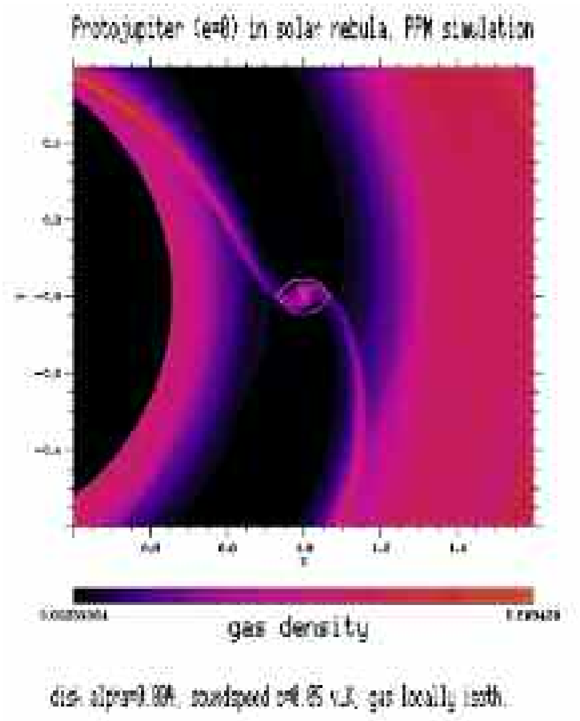


Fig. 2. A simulation of the formation of a planet in a protoplanetary disc, resulting from accretion of matter. From Ref.[9].

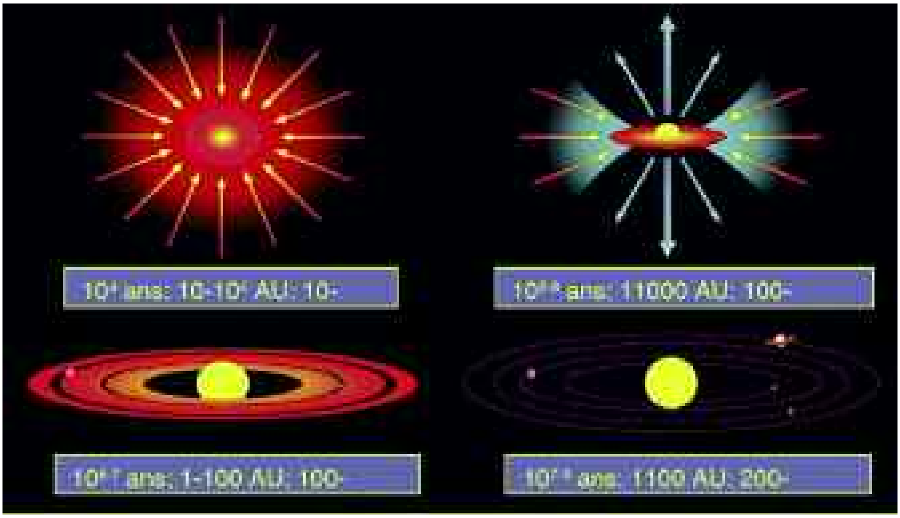


Fig. 3. The various stages of star and planet formation shown schematically, as they are now understood. Duration of each phase is in years, scale is in astronomical units (AU), and the temperature of the accreting gas in Kelvin.

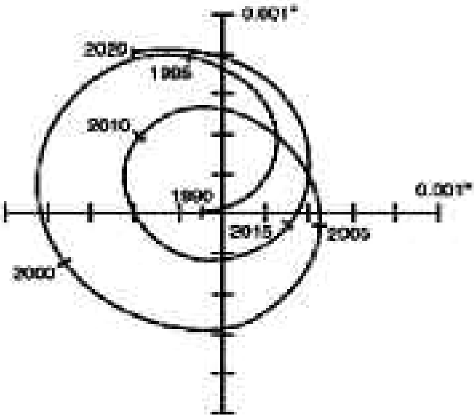


Fig. 4. The proposal of Michael Shao to detect the planet existence by measuring the reflex motion of its star during the orbit. This simulation shows the apparent motion the Sun would have on the sky, if observed 10 pc away, due to the presence of its planets (mostly Jupiter and Saturn). The amplitude is less than 1 milliarcsec, hence can only be detected with the angular resolution of an optical interferometer, located either in space or on the surface of the Earth.

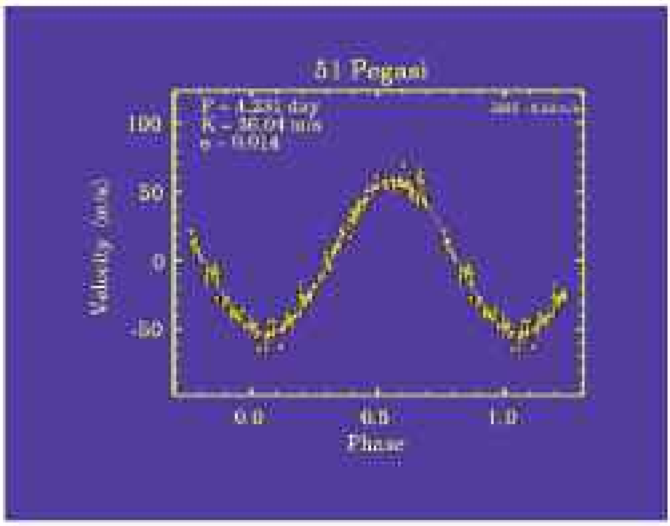


Fig. 5. The detection of the first planet by the induced motion of its star 51 Pegasi. The apparent velocity of the star in the direction of the observer is modulated with time (here labeled 'phase') at the rate of the orbiting period of 4.23 days (Source: M. Mayor, Geneva Observatory).

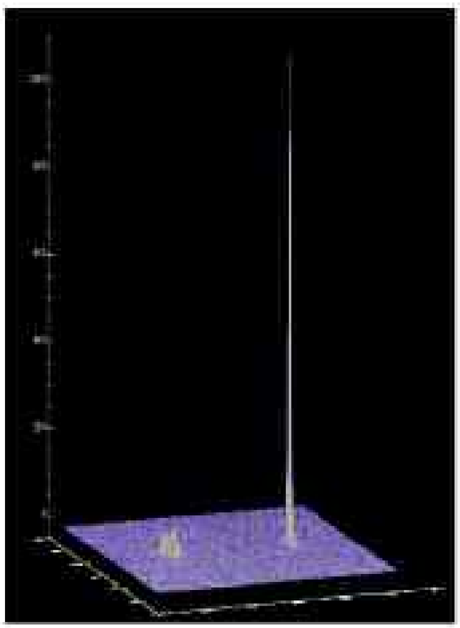


Fig. 6. The extraordinary performance of an adaptive optics correcting system on a large telescope, here on the European *Very Large Telescope*. The graph shows the distribution of the intensity of light at the focal plane of the telescope without (left) and with (right) the correction: the concentration is increased by more than one order of magnitude, allowing the telescope to practically reach its diffraction limit. This instrument is used for the observation shown on Fig. 7. (Source: ESO).

Fig. 7. The *direct* detection of a planetary mass object, in orbit around the brown dwarf 2M1207 at a distance of 55 astronomical units. The objects are situated 70 pc away from Earth, and are observed at the near-infrared wavelengths by the VLT telescope Yepun, equipped with adaptive optics (Ref.[16]). The second graph shows the planet position in comparison with the Solar system. (Source: A.-M. Lagrange & G. Chauvin, CNRS-INSU and ESO).

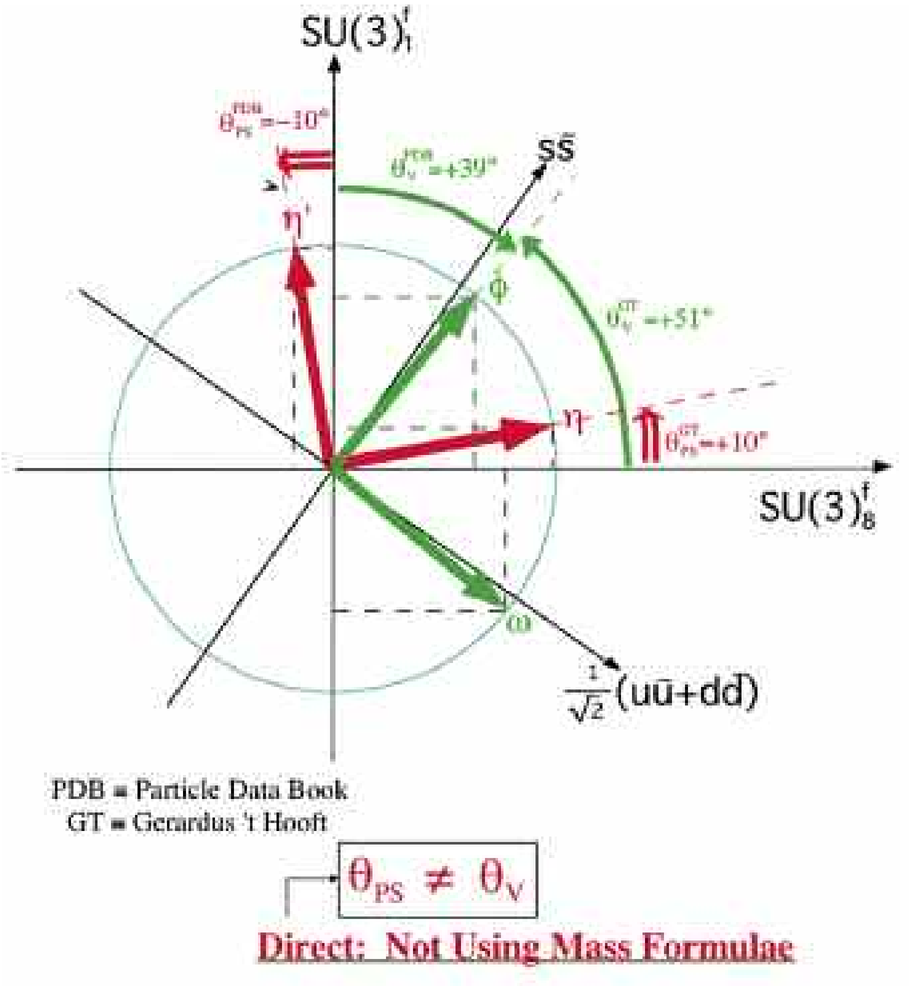

SU(3)$^f_1$
θ$_{PS}^{PDB}$=−10°
θ$_V^{PDB}$=+39°
ss̄
η'
φ
θ$_V^{GT}$=+51°
η
θ$_{PS}^{GT}$=+10°
SU(3)$^f_8$
ω
$\frac{1}{\sqrt{2}}$(uū+dd̄)
PDB = Particle Data Book
GT = Gerardus 't Hooft
θ$_{PS}$ ≠ θ$_V$
Direct: Not Using Mass Formulae

Figure 12.

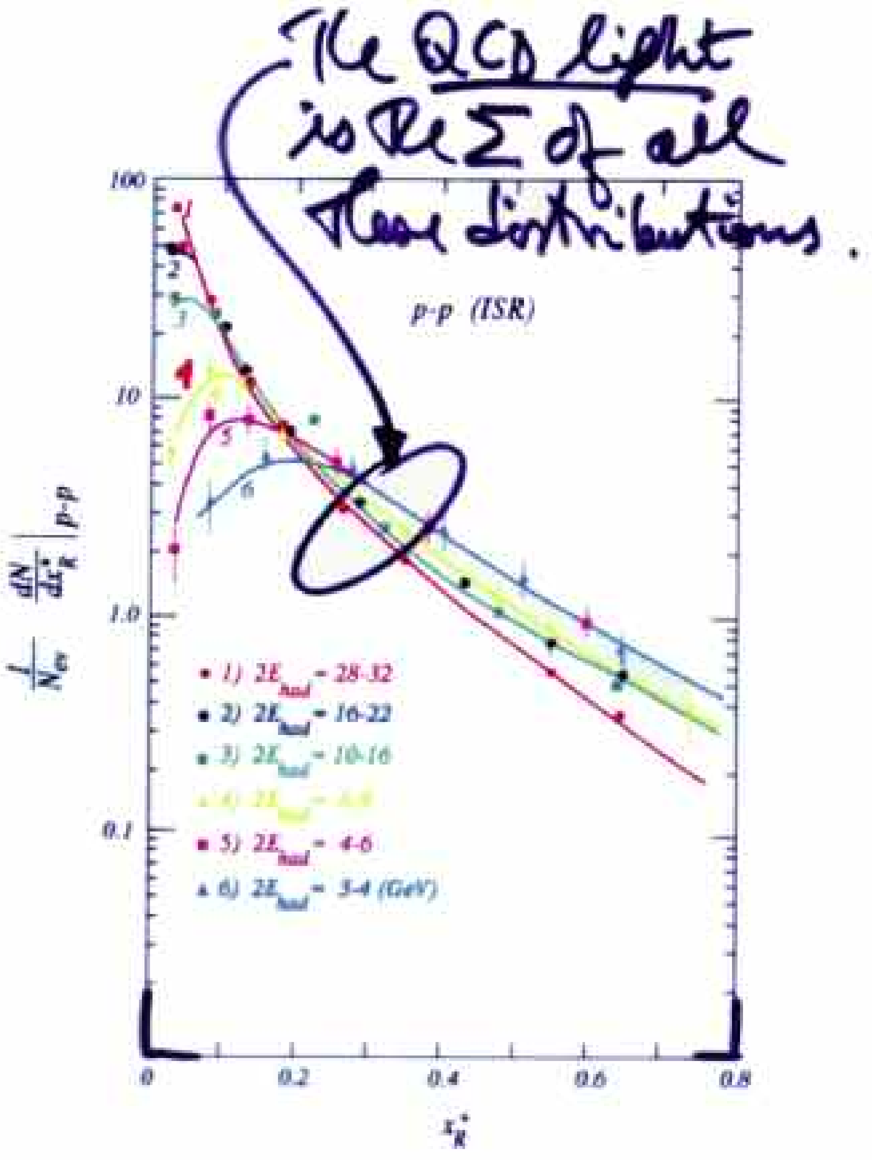

The QCD light is the Σ of all these distributions.
p-p (ISR)
$\frac{1}{N_{ev}} \frac{dN}{dx_R^*}\Big|_{p-p}$
1) $2E_{had}$ = 28-32
2) $2E_{had}$ = 16-22
3) $2E_{had}$ = 10-16
5) $2E_{had}$ = 4-6
6) $2E_{had}$ = 3-4 (GeV)
x_R^*

Figure 14.

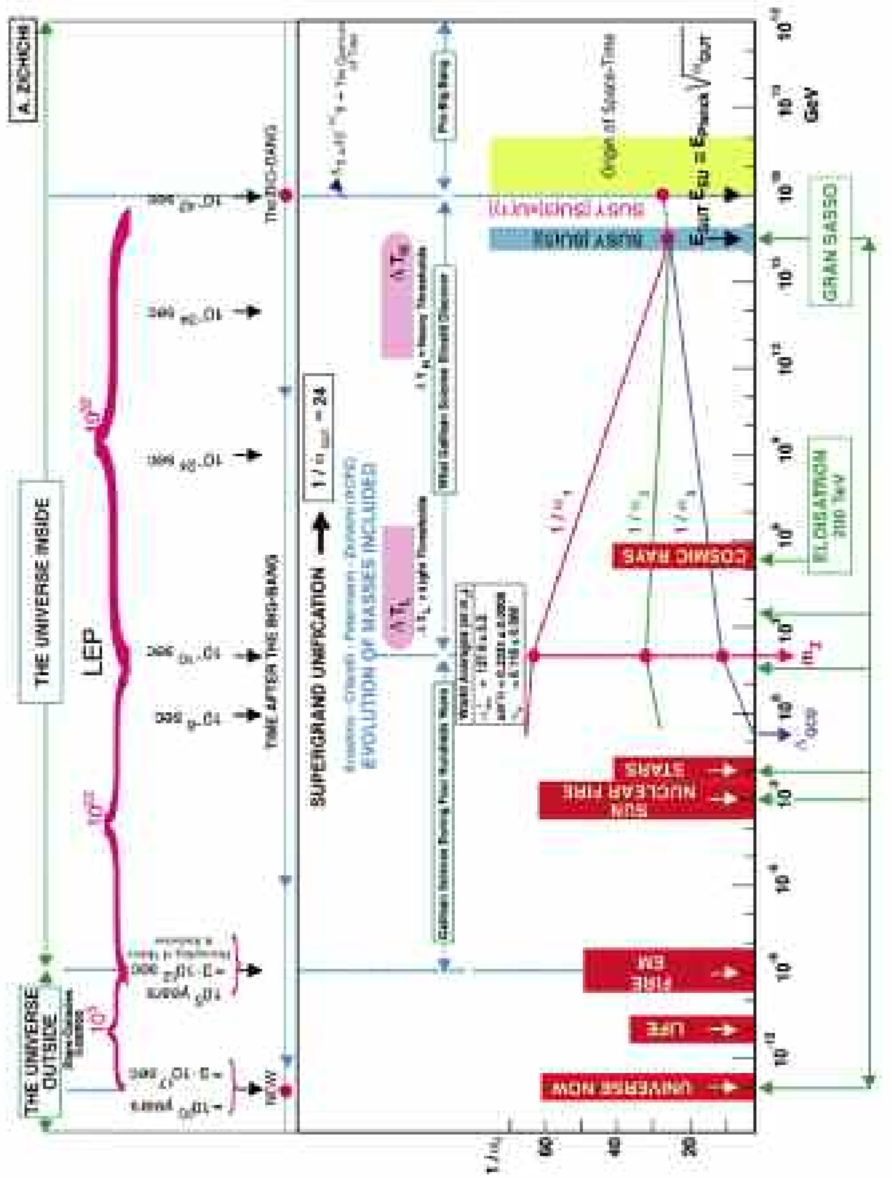

A. ZICHICHI
THE UNIVERSE INSIDE
THE UNIVERSE OUTSIDE
LEP
TIME AFTER THE BIG-BANG
SUPERGRAND UNIFICATION
EVOLUTION OF MASSES INCLUDED
UNIVERSE NOW
LIFE
FIRE EM
SUN NUCLEAR FIRE
STARS
COSMIC RAYS
SUSY (SU(5)xU(1))
Origin of Space-Time
GRAN SASSO
ELOISATRON
GeV

Figure 16.

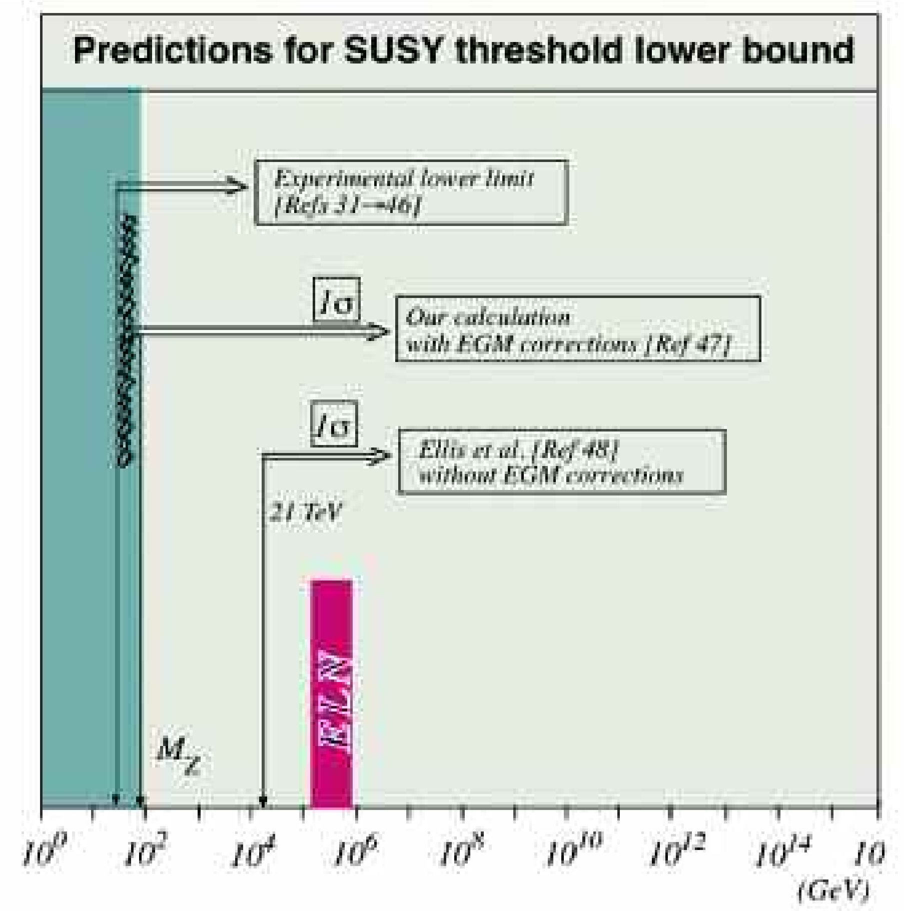
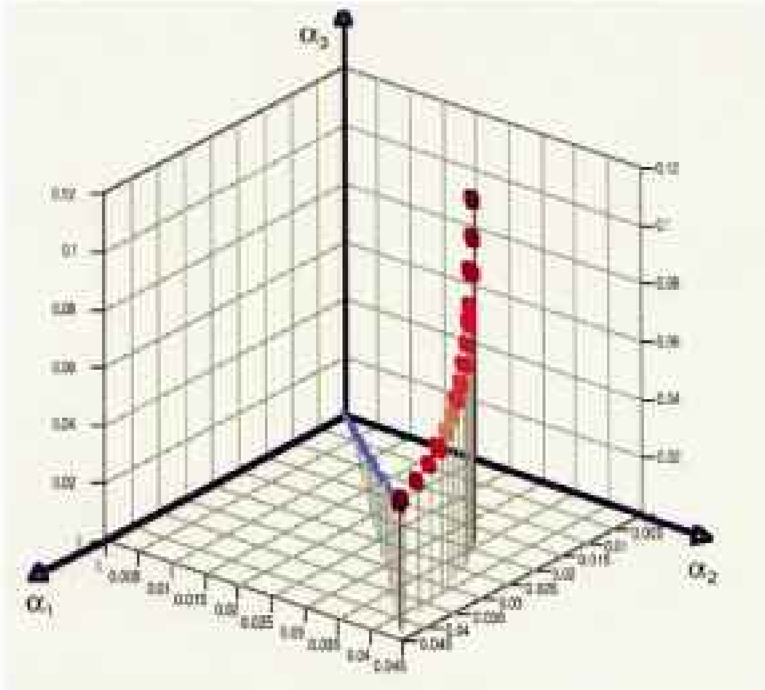

Predictions for SUSY threshold lower bound
Experimental lower limit
[Refs 31→46]
1σ
Our calculation
with EGM corrections [Ref 47]
1σ
Ellis et al. [Ref 48]
without EGM corrections
21 TeV
ELN
M_Z
10^0 10^2 10^4 10^6 10^8 10^{10} 10^{12} 10^{14} 10
(GeV)

Figure 17.



The points have a sequence of 100 GeV in energy. The last point where the 'ideal' platonic straight line intercepts the theoretical prediction is at the energy of the Grand Unification. This corresponds to $E_{GU} = 10^{16.2}$ GeV. Other detailed information on the theoretical inputs: the number of fermionic families, N_F, is 3; the number of Higgs particles, N_H, is 2. The input values of the gauge couplings at the Z^0-mass is $\alpha_3 (M_Z) = 0.118 \pm 0.008$; the other input is the ratio of weak and electromagnetic couplings also measured at the Z^0-mass value: $\sin^2 \theta_W (M_Z) = 0.2334 \pm 0.0008$.

Figure 19.

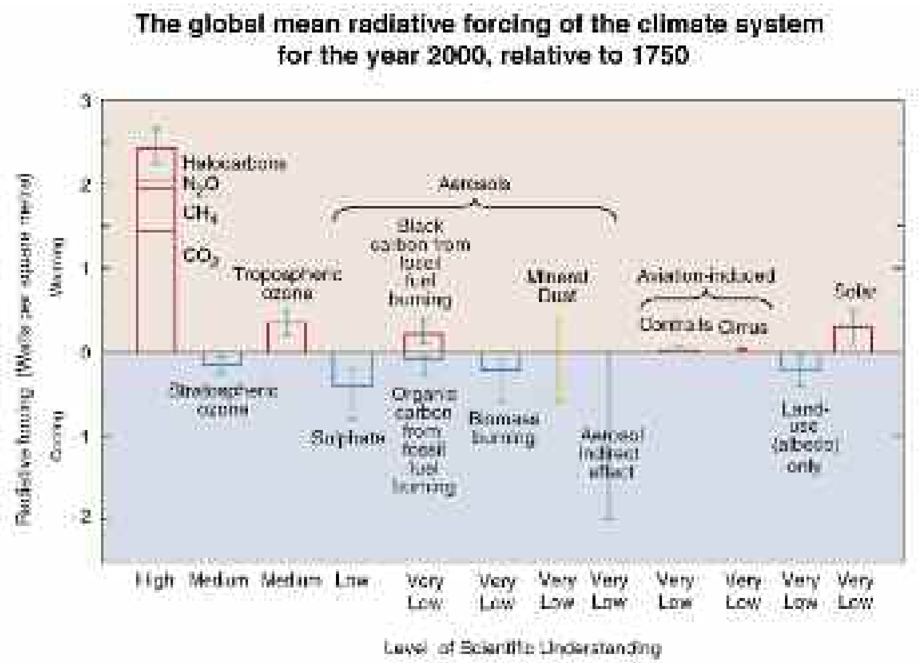


Figure 4.

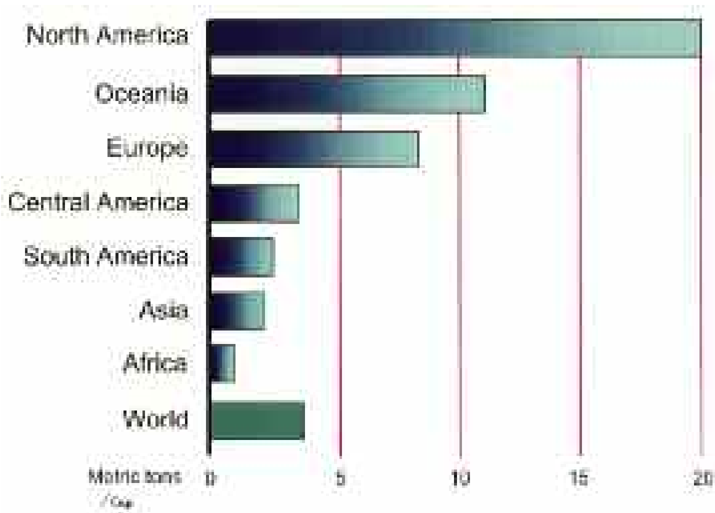


Figure 8.

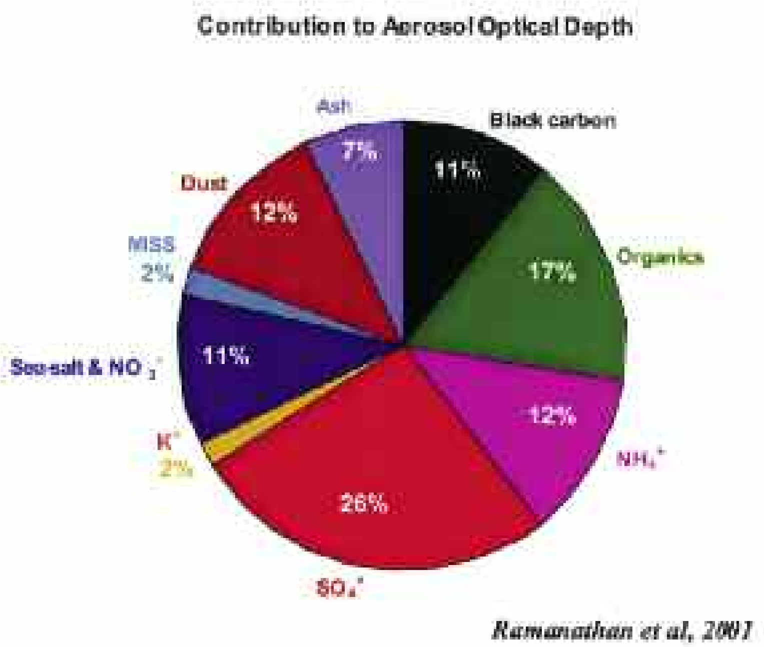


Figure 10.

Figure 11.

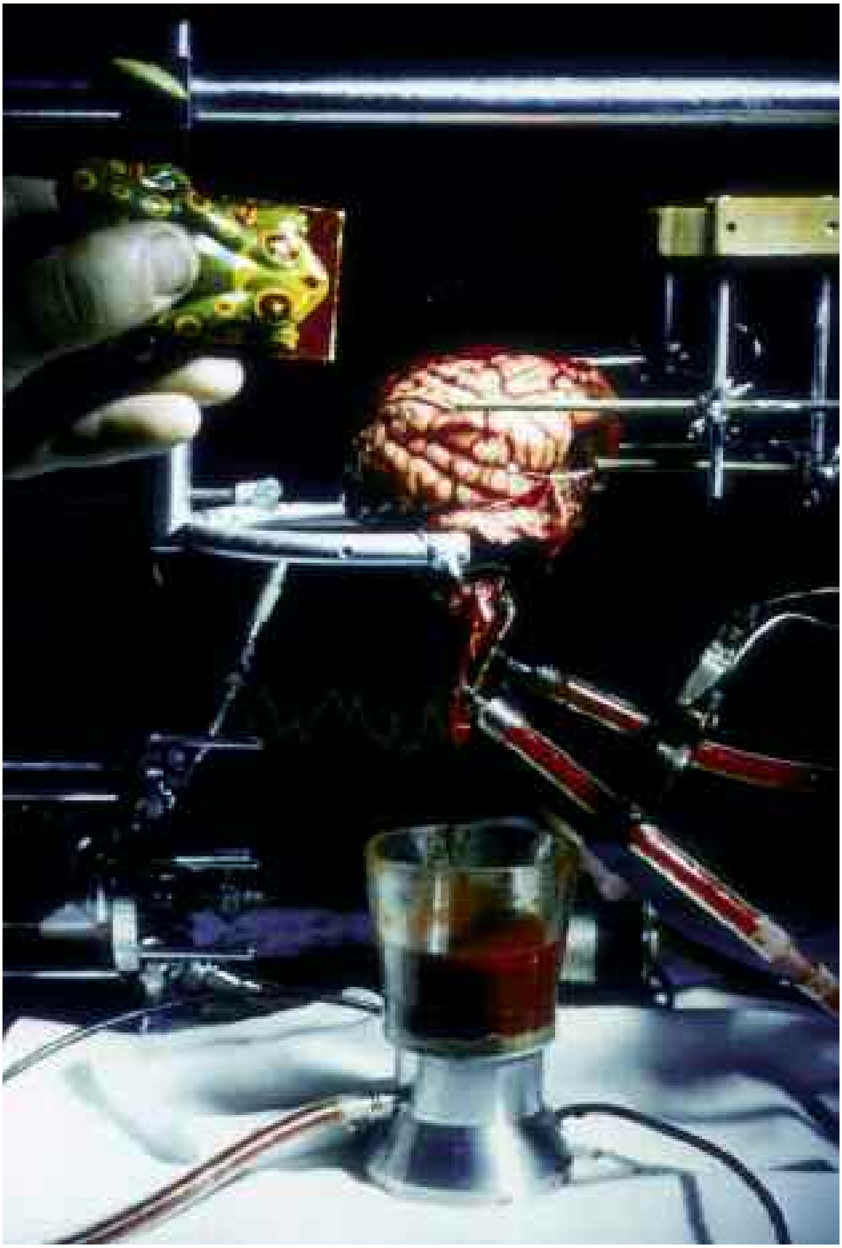


Figure 2. Isolated monkey brain without reservoir on mechanical perfusion. The preparation can record 'clicking' sound in the cortical auditory areas of the brain.

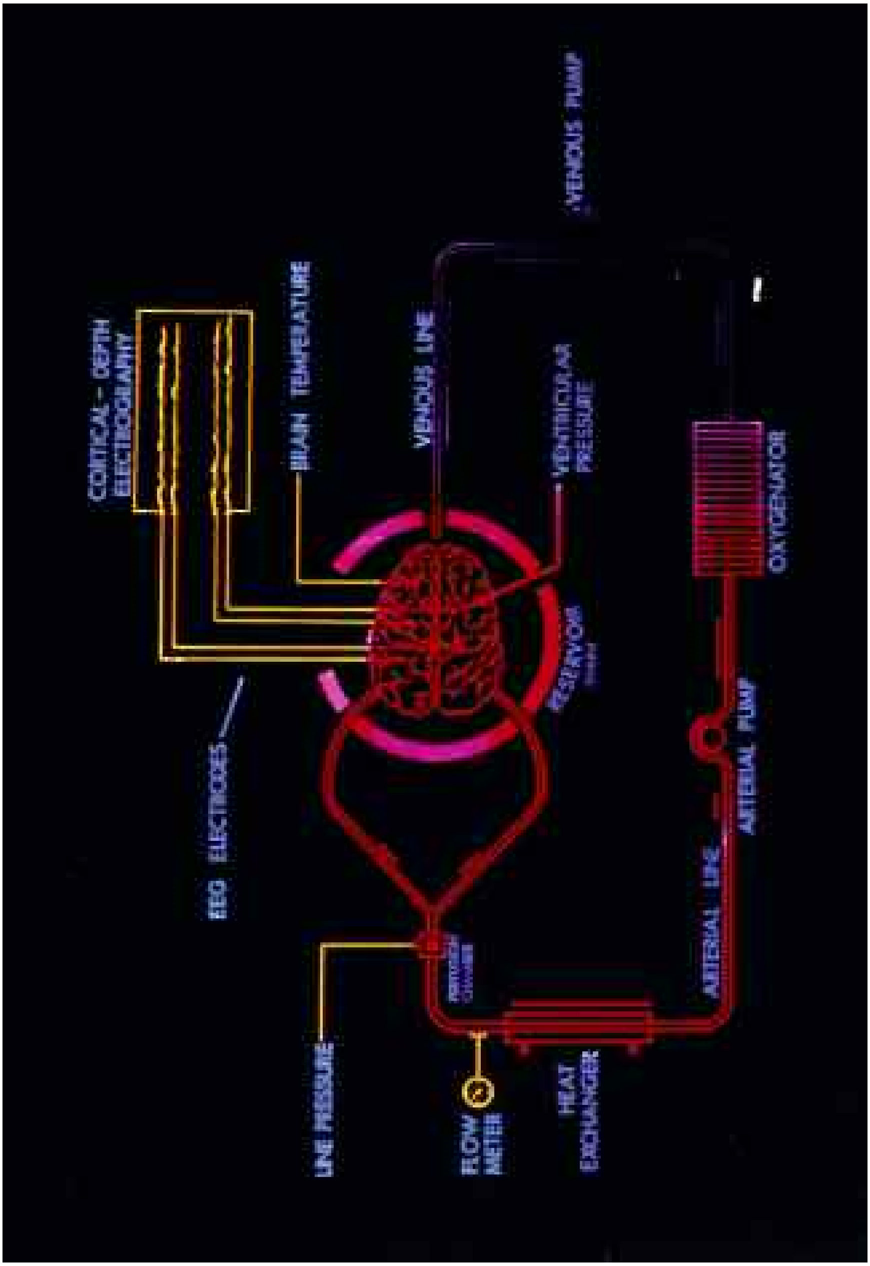


Figure 3. Diagram of the artificial perfusion system to viably support the isolated monkey brain.

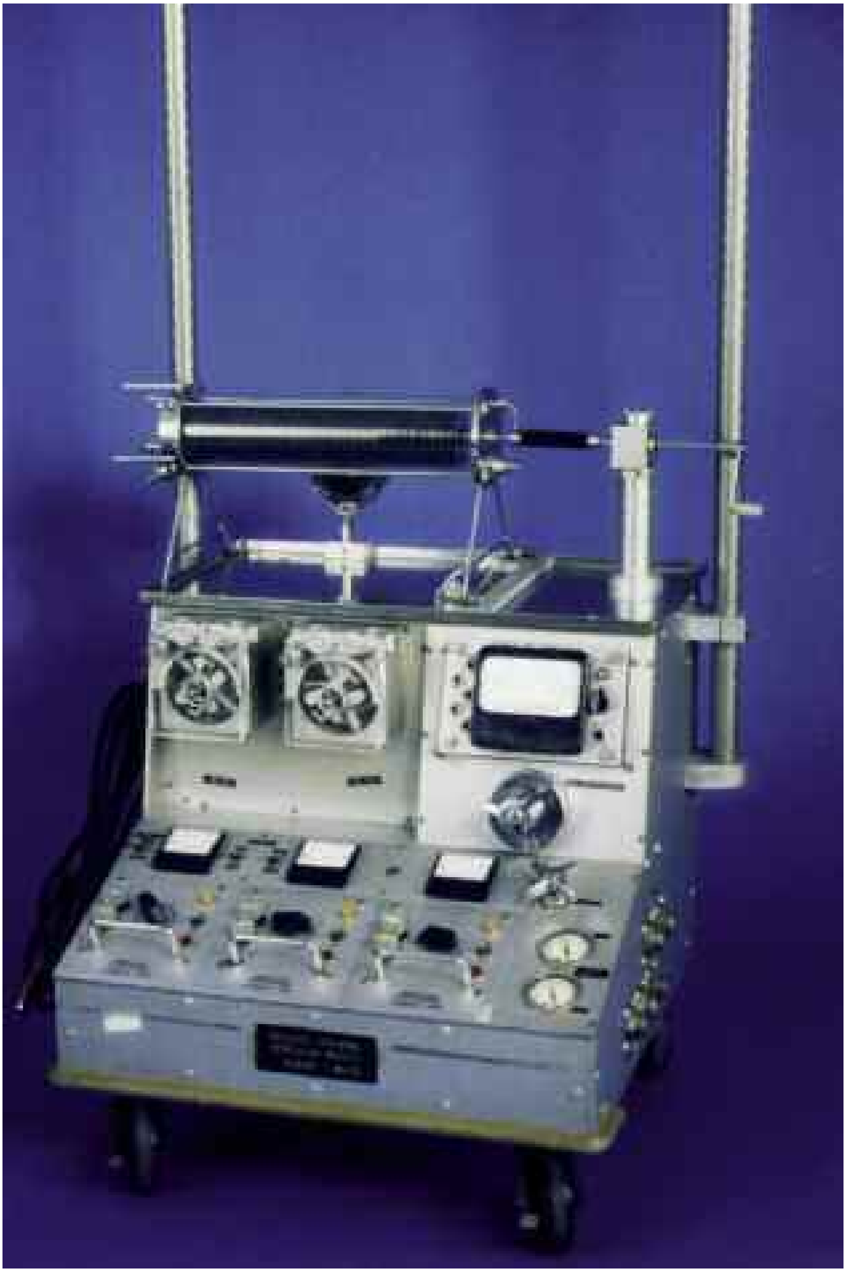


Figure 4. Mechanical heart-lung machine to keep human brain alive.

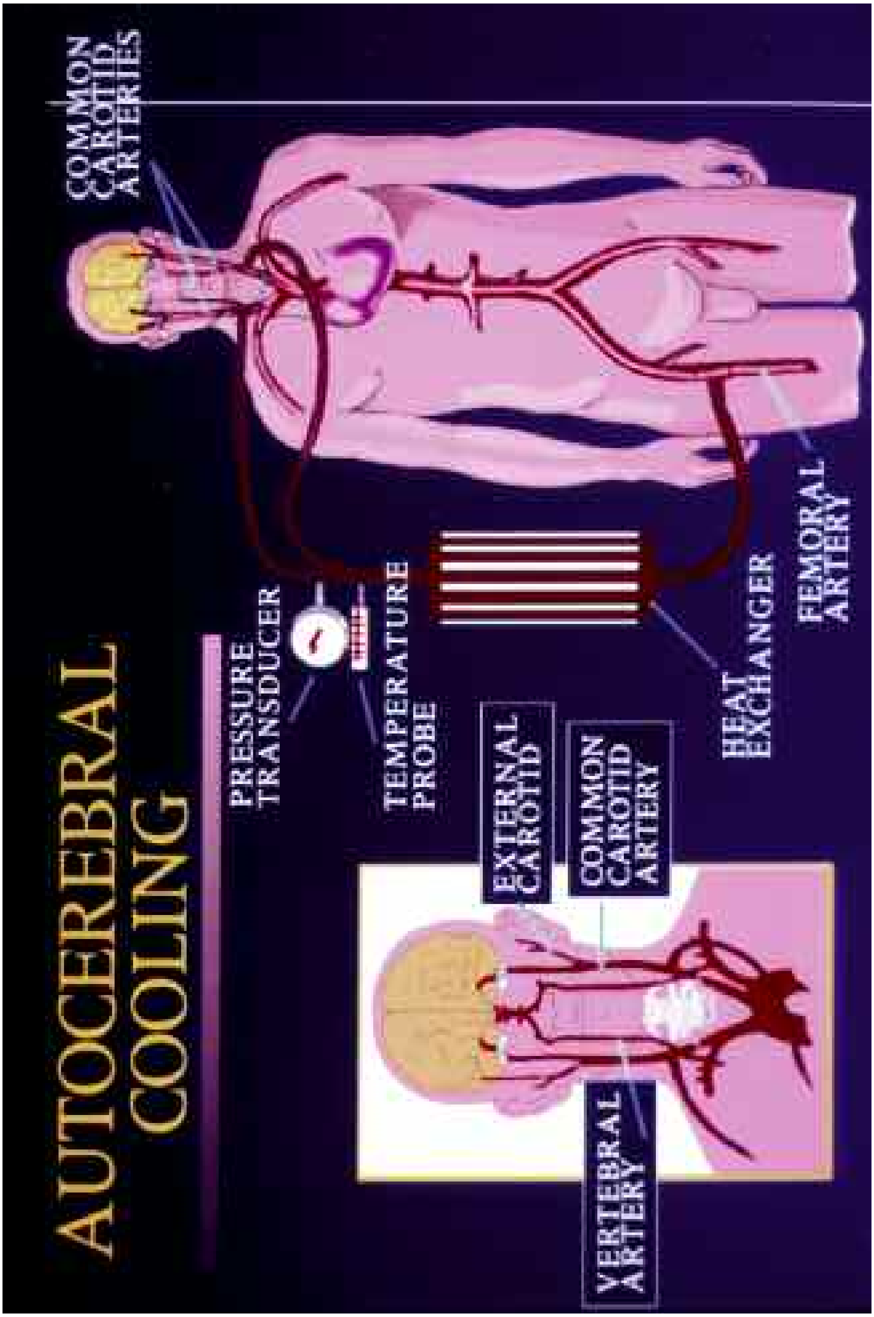


Figure 5. Autocerebral cooling – Patient cools his own brain without pump or oxygenation

Printed by
Tipolitografia SPEDIM
Via Serranti, 137
00040 Montecompatri, Rome
www.spedim.it

January 2006