Contemporary Perspectives in Philosophy and Methodology of Science

Edited by WENCESLAO J. GONZALEZ and JESUS ALCOLEA

netbiblo
www.netbiblo.com

General Editor

Wenceslao J. Gonzalez

CONTEMPORARY PERSPECTIVES IN PHILOSOPHY AND METHODOLOGY OF SCIENCE

ISBN 13: 978-0-9729892-3-7
ISBN 10: 0-9729892-3-4

Printed and bound by Gesbiblo, S.L.

Front cover: Gesbiblo, S.L.

First Published 2006 by Netbiblo, S.L.

Contents

PART III: FOCAL PHILOSOPHICAL PROBLEMS
IN EMPIRICAL AND FORMAL SCIENCES

About the Contributors to this Volume

Wenceslao J. Gonzalez is Professor of Logic and Philosophy of Science at the University of A Coruña. He has been Vice-dean of the School of Humanities and President of the Committee of Doctoral Programs at the University. He has been a visiting researcher at the Universities of St. Andrews, Münster and London (London School of Economics), as well as *Visiting fellow* at the Center for Philosophy of Science, University of Pittsburgh. He has given lectures at the Universities of Pittsburgh, Stanford, Quebec and Helsinki. The conferences in which he has participated include those organized by the Universities of Uppsala (Sweden), New South Wales (Australia), Bologna (Italy) and Canterbury (New Zealand).

After the monograph *La Teoría de la Referencia* (1986), Gonzalez has edited 21 volumes. He is the editor of the book *Science, Technology and Society: A Philosophical Perspective* (2005), and the monographic issues of journals on *Philosophy and Methodology of Economics* (1998) and *Lakatos's Philosophy Today* (2001). He has published 75 papers which include "Economic Prediction and Human Activity" (1994), "On the Theoretical Basis of Prediction in Economics" (1996), "Rationality in Economics and Scientific Predictions" (1997), "Lakatos's Approach on Prediction and Novel Facts" (2001), "Rationality in Experimental Economics" (2003), "From *Erklären-Verstehen* to *Prediction-Understanding*: The Methodological Framework in Economics" (2003), "The Many Faces of Popper's Methodological Approach to Prediction" (2004), and "The Philosophical Approach to Science, Technology and Society" (2005).

Colin Howson is Professor at the *London School of Economics*, where he is the Convenor of the Department of Philosophy, Logic and Scientific Method. He has been President of the British Society for the Philosophy of Science. He is co-author of the influential book *Scientific Reasoning: The Bayesian Approach* (1989; 3rd ed., 2006) and author of the monograph *Hume's Problem* (2000). Previously he was editor of the volume *Method and Appraisal in the Physical Sciences* (1976).

Howson has published many papers on Bayesianism as well as criticizing Popper's views on probability: "Must the Logical Probability of Laws Be Zero?" (1973); "The Rule of Succession, Inductive Logic and Probability Logic" (1975); "The Prehistory of Chance" (1978); "Miller's So-Called Paradox of Information" (1979); "Methodology in Non-Empirical Disciplines" (1979); "Bayesianism and Support by Novel Facts" (1984); "Popper's Solution of the Problem of Induction" (1984); "Popper, Prior Probabilities and Inductive Inference" (1987); "Accommodation, Prediction and Bayesian Confirmation Theory" (1989); "Fitting your theory to the facts: Probably not such a bad thing after all" (1990); "The 'Old Evidence' Problem" (1991); "Bayesian Conditionalization and Probability Kinematics" (1994); "Theories of Probability" (1995); "Probability and Logic" (2001); "Bayesianism in Statistics" (2002); and "Bayesian Evidence" (2003).

Donald A. Gillies is Professor at University College London. He has been President of the British Society for the Philosophy of Science. From 1982 to 1985 he edited the *British Journal for the Philosophy of Science*. Previously, he studied mathematics and philosophy at Cambridge, and then began graduate studies in Professor Sir Karl Popper's department at the London School of Economics. After completing his Ph. D. on the foundations of probability with Professor Imre Lakatos as supervisor, he joined the staff of the University of London, where he was Professor of the Philosophy of Science and Mathematics at King's College London.

Gillies has published extensively in philosophy of science and mathematics. He is the author of the monographs *An Objective Theory of Probability* (1973), *Frege, Dedekind and Peano on the Foundations of Arithmetic* (1982), *Philosophy of Science in the Twentieth Century. Four Central Themes* (1993), *Artificial Intelligence and Scientific Method* (1996), and *Philosophical Theories of Probability* (2000). He is the editor of the volume *Revolutions in Mathematics* (1992). Among his papers are "Dynamic Interactions with the Philosophy of Mathematics" (2001), "El problema de la demarcación y la Medicina alternativa" (2004), and "El problema de la inducción y la Inteligencia Artificial" (2004).

John Worrall is Professor at the *London School of Economics*, where he has been Convenor of the Department of Philosophy, Logic and Scientific Method. He was the editor of the *British Journal for the Philosophy of Science* (1974-1983). He is co-director of the *Center for Philosophy of Natural and Social Science* at LSE and member of Imre Lakatos Memorial Committee (1974-). He has been co-editor of Lakatos's works: *Proofs and Refutations* (1976), *The Methodology of Scientific Research Programmes* (1978) and *Mathematics, Science and Epistemology* (1978). He has also edited *The Ontology of Science* (1994).

Worrall has paid particular attention to history of science: "Fresnel, Poisson and the White Spot: The role of successful predictions in the acceptance of scientific theories" (1989), and "Prediction and the Periodic Table" (2001). His research is generally focus on central topics on philosophy and methodology of science: "Imre Lakatos: Philosopher of Mathematics and Philosopher of Science" (1976), "The Ways in which the Methodology of Scientific Research Programmes improves on Popper's Methodology" (1978), "Scientific Realism and Scientific Change" (1982), "Rationality, Sociology and Symmetry Thesis" (1990), "Structural Realism: the Best of Both Worlds?" (1996), "De la Matemática a la Ciencia: Continuidad y discontinuidad en el Pensamiento de Imre Lakatos" (2001), "Programas de investigación y heurística positiva: Avance respecto de Lakatos" (2001), "*What* Evidence in Evidence-Based Medicine?," (2002), and "Normal Science and Dogmatism, Paradigms and Progress: Kuhn 'versus' Popper and Lakatos" (2003).

Peter J. Clark is Professor of the History and Philosophy of Science in the University of St Andrews (United Kingdom), where he is Head of the School of Philosophical and Anthropological Studies. He was the Editor of *The British Journal for the Philosophy of Science* from 1997 until 2004. His dissertation supervisors at the London School of Economics, where he took his Ph. D., were Imre Lakatos and Elie Zahar.

Clark's main research interests lie in the history and philosophy of natural science and mathematics. He is co-editor of the books *Mind, Psychoanalysis and Science* (1988), *Reading Putnam* (1994), and *Philosophy of Science Today* (2003). Among his writings are "Determinism and Probability in Physics" (1987), "Logicism, the Continuum, and Anti-Realism" (1993), "Poincaré, Richard's Paradox and Indefinite Extensibility" (1995), "Popper on Determinism" (1995), "Statistical Mechanics and the Propensity Interpretation of Probability" (2001), and "Frege, Neo-Logicism and the Good Company Objection" (2003).

Franz M. Wuketits is Professor at the University of Vienna. His main fields of research are evolutionary epistemology and evolutionary ethics. He studied zoology, paleontology and philosophy. He presented his Dissertation in 1978. His academic positions have been at the University of Vienna, University of Graz and University of Technologyof Vienna. He belongs to the board of directors of the Konrad Lorenz Institute for evolution and cognition research.

Wuketits has published 30 books. He is the editor or co-editor of 12 volumes and has written more than 300 papers. Among his books are *Evolutionary Epistemology and Its Implications for Humankind* (1990), *Die Entdeckung des Verhaltens: Eine Geschichte der Verhaltensforschung* (1995), *Naturkatastrophe Mensch: Evolution ohne Fortschritt* (1998), *Warum uns das Boese fasziniert: Die Natur des Boesen und die Illusionen der Moral* (1999), *Ausgerottet- ausgestorben: Ueber den Untergang von Arten, Voelkern und Sprachen* (2003), *Handbook of Evolution. Vol. 1: The Evolution of Human Societies and Cultures* (2003), co-edited with Christoph Antweiler, and *Handbook of Evolution. Vol. 2: The Evolution of Living Systems* (2005), co-edited with Francisco J. Ayala.

Emilio Muñoz is Research Professor at the Higher Council of Scientific Research (CSIC) in Madrid. He has been the President of this scientific institution. He is Head of the Department of Science, Technology and Society. He is a member of the European Molecular Biology Organization, EMBO. He was President of the Committee of Biotechnology of the CEFI Foundation and of the *European Interuniversity Association on Society, Science and Technology* (ESST).

Muñoz has been working on molecular biology and biochemistry as well as in public policy of science and technology. Among his publications are the books: *Genes para cenar: Biotecnología y las nuevas especies* (1991), *Una visión de la Biotecnología: Principios políticas y problemas* (1994), *Biotecnología, industria y sociedad: el caso español* (1997), *El establecimiento de la Bioquímica y Biología Molecular en España* (1997), *Cuarenta años de la Sociedad Española de Bioquímica y Biología Molecular* (2004), and *Radiografía de la Investigación pública en España*, co-edited with Jesús Sebastián (2006). Among the papers can be pointed out "Nueva biotecnología y sector agropecuario: El reto de las racionalidades contrapuestas" (1998), "La investigación en la España de hoy: Mapa de acciones y constricciones y su reflejo en el paisaje de 1997" (1998), "Radioisótopos y nueva Biología: Una innovación metodológica con repercusiones sociales" (2003), and "La investigación biomédica en España y sus circunstancias" (2003).

Anna Estany Profitos is Professor of Logic and Philosophy of Science at the Autonomous University of Barcelona and Head of the Department of Philosophy. She was awarded her *Master of Arts* by the University of Indiana (1985) and has researched at the University of California at San Diego.

Her books are *Modelos de cambio científico* (1990), *Introducción a la Filosofía de la Ciencia* (1993), *Vida, muerte y resurrección de la conciencia. Análisis filosófico de las revoluciones científicas en la psicología contemporánea* (1999), and *La fascinación por el saber* (2001). Estany is co-author of *Manual de prácticas de Filosofía de la Ciencia* (2000), and *¿Eureka? El trasfondo de un descubrimiento sobre el cáncer y la Genética Molecular* (2003). Among her papers are "Reconstrucción de casos históricos a partir del modelo de progreso científico de L. Laudan" (1998), "Thomas Kuhn: Hacia una metateoría unificada de la Ciencia" (1998), "Ventajas epistémicas de la cognición socialmente distribuida" (2001), "The Theory-laden Thesis of Observation in the Light of Cognitive Psychology" (2001), and "Progress and Social Impact in Design Sciences" (2005).

Peter K. Machamer is Professor at the University of Pittsburgh, where he was Director of the Department of History and Philosophy of Science for 15 years (1978-1993). He is also a member of the Center for the Study of the Neural Basis of Cognition. Previously, he has taught in Illinois Institute of Technology (Chicago), Queens College (N. York), and Ohio State University.

Machamer is co-editor of *Perception: Historical and Philosophical Studies* (1978), *Proceedings of the Philosophy of Science Association* (1986), *A Companion to Galileo* (1998), *Scientific Controversies* (2000), *Theory and Method in the Neurosciences* (2001), and *The Blackwell Guide to the Philosophy of Science* (2002). He is co-author of *Introduction to the Philosophy of Science* (1992). He is the author of "Feyerabend and Galileo: The Interaction of Theories and the Reinterpretation of Experience" (1973), "Galileo and the Causes" (1978), "Galileo, Mathematics and Mechanism" (1998), "The Nature of Metaphor and Scientific Descriptions" (2000), "Rational Reconstructions Revised" (2001), "El éxito de Kuhn, 40 años después" (2004) and "Las revoluciones de Kuhn y la Historia 'real' de la Ciencia: El caso de la revolución galileana" (2004).

Jose Sanmartin is Professor of Logic and Philosophy of Science at the University of Valencia and Director of *Centro Reina Sofía para el estudio de la violencia*. He has been researcher of *Alexander von Humboldt Foundation* (1974-1976, 1989) and visiting researcher at the *Max Planck Institute* of Human Ethology (1987, 1988 and 1989). He has been Vice-president of the *European Association on Society, Science and Technology* and President of *The Society for Philosophy and Technology* (1995-1997).

Sanmartin has published the monographs *Una introducción constructiva a la teoría de modelos* (1978), *Filosofía de la Ciencia.* (1982), *Los Nuevos Redentores* (1987), *Tecnología y futuro humano* (1990), *La violencia y sus claves* (2000), *La mente de los violentos* (2002), and *El terrorista. Cómo es. Cómo se hace* (2005). He is editor of *Genética y conducta* (1986), *Ciencia, Tecnología y Sociedad* (1990), *Superando fronteras* (1994), *New Worlds, New Technologies, New Issues* (1992), *Estudios sobre sociedad y tecnología* (1992), *Violence*: *from Biology to Society* (1997), *Violencia, televisión y cine*. (1998), *Ética y televisión* (1998), *Violencia contra niños* (1999), *Violencia* (2000), *Violence and Psychopathy* (2001), and *El laberinto de la violencia* (2004).

Javier de Lorenzo is Professor of Logic and Philosophy of Science at the University of Valladolid. His main research interests lie in the history and philosophy of mathematics.

He is the author of the monographs *Introducción al estilo matemático* (1971), *La Filosofía de la Matemática de Jules Henri Poincaré* (1974), *El método axiomático y sus creencias* (1977), *Kant y la Matemática. El uso constructivo de la razón pura* (1992), *La Matemática: De sus fundamentos y crisis* (1998), *El infinito matemático* (2001), and *Filosofías de la Matemática: fin de siglo XX* (2001). He is co-editor of *Calculemos.*

Matemáticas y Libertad (1996) and editor of the two volumes of *Obras escogidas. Miguel Sánchez Mazas* (2002 and 2003). Among his papers are “Matemáticas y crítica” (1981), “Historia de la Matemática: Problemática actual” (1995), and “El ‘programa Poincaré’ o funciones del matemático” (2004).

Jesus Alcolea is Reader [Titular Professor] of Logic and Philosophy of Science at the University of Valencia. He has been Director of the Departament of Logic and Philosophy of Science, and he is currently the Vicedean of the School of Philosophy and Science of the Education. He undertook research at Smith College (Northampton, Massachusetts) with Thomas Tymoczko.

From the beginning, Alcolea’s focus of research has been on philosophy of mathematics. He is the author of the book *Logicismo, Formalismo, Intuicionismo* (1985) as well as many papers in this field: “Proof as a Way to Increase Understanding” (1996), “Metaphysical Myths, Mathematical Practice: The Ontology and Epistemology of the Exact Sciences” (1996), “Demostración como comunicación” (1997), “Fallibilism and Realism in Lakatos” (1999), “Vigencia del Pensamiento filosófico-matemático de Imre Lakatos” (2001), “Los conceptos matemáticos en el Mundo 3 de Popper” (2004); and “La extensión de los conceptos matemáticos” (2005).

Novelty and Continuity in Philosophy and Methodology of Science

Wenceslao J. Gonzalez

Nowadays, philosophy and methodology of science appear as a combination of novelty and continuity. This blend is clear both in the general approaches to science, those thought of as any science (mostly empirical sciences: natural, social, and artificial), and in the specific perspectives on every science, either formal (mathematics...) or empirical. Moreover, it also happens that some innovations come from the special analysis (e.g., philosophy of biology or philosophy of the sciences of the artificial)[1] to reach later general philosophy and methodology of science.

On the one hand, there are *new topics* for philosophical reflection, such as key issues in philosophy of medicine and central problems raised by neuroscience. Thus, new contents have brought attention to aspects that previously went almost unnoticed (e.g., methodological problems on clinical trials or epistemological issues on the use of designs in science). In addition, there are new angles for philosophical study, such as the repercussion of society on scientific activity (in aims, processes, and results).[2] But, on the other hand, the *background* of the main philosophical and methodological trends of the twentieth century is, in many ways, still in place.[3] Moreover, there has been no clear breakthrough in this philosophical domain in recent decades. *De facto*, the present panorama can be described as a combination of different philosophic-methodological orientations that coexist as wide-ranging frameworks for the research lines.

Following these features, this paper considers several aspects: 1) the main trends in general philosophy and methodology of science, which include those positions that were dominant before the early 80's and the new tendencies ("naturalistic turn," "social turn," and the realist conceptualizations); 2) the naturalistic approaches, both in the case of the "cognitive turn" and in the perspective of normative naturalism; 3) the social concern on science; 4) the interest in "scientific realism," taking into account central views on scientific realism and some influential conceptions build up on realist grounds; and 5) the increased emphasis on visions based on theories of probability, where a key role is played by Bayesianism. Afterwards, there are details on the structure and origin of the present volume as well as a bibliography on the topics considered in this initial chapter.

1. Main Trends in General Philosophy and Methodology of Science

There are some orientations in general philosophy and methodology of science that have been influential. They conform the main trends to date. For several decades —from the

---

1 Recent developments on "Design sciences" are connected to the analysis made in Simon, H. A., *The Sciences of the Artificial,* 3rd ed., The MIT Press, Cambridge, MA, 1996. In addition, new contributions on what "applied science" is are based on that analysis, cf. Gonzalez, W. J. (ed.), *Racionalidad, historicidad y predicción en Herbert A. Simon,* Netbiblo, A Coruña, 2003.

2 Cf. Kitcher, Ph., *Science, Truth, and Democracy,* Oxford University Press, Oxford, 2001; and Gonzalez, W. J. (ed.), S*cience, Technology and Society: A Philosophical Perspective,* Netbiblo, A Coruña, 2005.

3 Among the topics of that background are those analyzed in Gillies, D. A., *Philosophy of Science in the Twentieth Century. Four Central Themes, B.* Blackwell, Oxford, 1993.

mid-20's to the late 70's—, the focus of attention lay mostly in three directions: firstly, in the different positions of the "received view";[4] secondly, in the successive versions of Karl Popper's approach;[5] and, thirdly, in the later "historical turn," where diverse conceptions of scientific change were offered. These perspectives based on the historicity of scientific activity were developed mostly in the 60's and the 70's. Key figures in this turn have been Thomas Kuhn, Imre Lakatos, and Larry Laudan. Although they highlight the relevance of history of science to philosophy of science, there are also relevant differences among their methodological approaches.[6]

Afterwards —since the early 80's to now—, the general philosophic-methodological discussion moved towards new main trends. Although they assumed the central claims of the "historical turn" (the relevance of scientific change, science as human activity in a social setting, etc.), the new positions have emphasized new aspects related to scientific activity. Usually, they have been built up around three possibilities: the naturalist approaches, the views based on social concern, or the conceptions with special emphasis on scientific realism.[7] In addition, there is an increased emphasis on views based on theories of probability, where Bayesianism has played a leading role.[8]

1.1. The Previous Dominant Trends

Within contemporary philosophy and methodology of science several main trends can be pointed out since the manifesto of the Vienna Circle (1929).[9] This event can be considered as a crucial point of this field, insofar as the constitutive elements of science (language, structure, knowledge, method, ...) were —in one way or another— analyzed in the philosophical movement that was consolidated with that famous statement. Furthermore, these neopositivist authors draw attention to some aspects that, *de facto*, put science in the center of the philosophical discussion of the twentieth century.

Following this trend is the "received view," which has its roots in that initial logical positivism of Moritz Schlick and Rudolf Carnap, and which gains a more sophisticated

---

4 "Received view" is an expression used by Hilary Putnam in the same year that Thomas Kuhn published his book on "scientific revolutions." Cf. Putnam, H., "What Theories Are Not," in Nagel, E., Suppes, P. and Tarski, A. (eds.), *Logic, Methodology and Philosophy of Science,* Stanford University Press, Stanford, 1962, pp. 240-251. That expression grasps the momentum of the methodology of science based on the verificationist tradition.

5 In Karl Popper's approach three successive versions can be distinguished, according to the main emphasis of his perspective: logico-methodological, epistemological, and ontological. Cf. Gonzalez, W. J., "La evolución del Pensamiento de Popper," in Gonzalez, W. J. (ed.), *Karl Popper: Revisión de su legado,* Unión Editorial, Madrid, 2004, pp. 23-194; especially, pp. 24-65.

6 On their conceptions, cf. Gonzalez, W. J. (ed.), *Análisis de Thomas Kuhn: Las revoluciones científicas*, Trotta, Madrid, 2004; Gonzalez, W. J. (ed.), *La Filosofía de Imre Lakatos: Evaluación de sus propuestas*, UNED, Madrid, 2001; and Gonzalez, W. J. (ed.), *El Pensamiento de L. Laudan. Relaciones entre Historia de la Ciencia y Filosofía de la Ciencia*, Publicaciones Universidad de A Coruña, A Coruña, 1998.

7 The list can be enlarged with other views, such as post-modern conceptions. In this regard, cf. Koertge, N. (ed.), *A House Built on Sand: Exposing Postmodern Myths about Science*, Oxford University Press, New York, 1998.

8 There are, of course, other general philosophic-methodological tendencies during this period. Among them is the Structuralist Program, which has been influential in some countries in Europe and South America. On this view, cf. Balzer, W., Moulines, C. U. and Sneed, J. D., *An Architectonic for Science. The Structuralist Program,* Reidel, Dordrecht, 1987.

9 Cf. Hahn, H., Carnap, R. and Neurath, O., *Wissenschaftliche Weltauffassung: Der Wiener Kreis,* Wolff, Vienna, 1929. Translated as: *The Scientific Conception of the World: The Vienna Circle,* Reidel, Dordrecht, 1973.

development through logical empiricism.[10] Thus, the original methodological insistence on "verification" moved towards "verifiability" and, later on, became only "confirmability."[11] These adaptations of the central claims of this philosophic-methodological tendency made them be very influential until the mid-60's. Moreover, these thinkers analyzed a large number of topics (explanation, prediction, theories, laws, ...) that remained later on in the posterior philosophic-methodological discussion. The subsequent modifications came from the variation in the *kind of approach* (historical, social, etc.), which led to different conclusions.

During all this period (from the late 20's to the mid-60's, and beyond), Karl Popper developed his views on science.[12] His positions in epistemology (critical rationalism) and methodology (falsificationism) were commonly seen as an alternative to the "received view." In his approach, the critical attitude towards scientific knowledge and the fallibilism regarding the results of science have a key role. His emphasis on the contents themselves of science, rather than on the agents that do science or on the social environment where scientific activity is developed, were later criticized in different ways (in the posterior *turns*: historical, naturalistic, social, ...). But Popper's stress on *objectivity* in scientific knowledge still remains —in one way or another— in authors in favor of scientific realism.

A neat turning point came with Kuhn's proposals. In effect, he has led the "historical turn" since his book of 1962.[13] Then the focus was put on science as *human activity*, and the previous emphasis on isolated theories was replaced by large "paradigms" (i.e., theory, law, application, and instrumentation) in order to understand scientific change. In the explicit revision of falsificationism, the central figure was Lakatos. He developed a direct criticism of Popper through the methodology of "scientific research programs." In this conception, he tried to combine historicity of science and objectivity of its contents. The progressive character of science was associated to prediction of novel facts.[14] Afterwards, many authors, such as Laudan, continued the "historical turn," where a close link was established between history of science and philosophy of science.[15]

It seems remarkable that the "historical turn" changed the philosophic-methodological panorama completely. The new emphasis was the idea of science as a human activity and, therefore, as a social event. The change in the *turn* includes historicity as an internal constituent of science as well as an external factor. Both elements combined —human activity and historicity— lead to another key element: the decision making of the scientific

10 In fact, in some ways, the "logical empiricism" is already in Hans Reichenbach. He has clear differences with Carnap in central topics, cf. GONZALEZ, W. J., "Reichenbach's Concept of Prediction," *International Studies in the Philosophy of Science,* v. 9, n. 1, (1995), pp. 37-58.

11 Cf. SUPPE, F. (ed.), *The Structure of Scientific Theories*, University of Illinois Press, Urbana, 1974 (2nd. ed. 1977).

12 He also has publications that have been influential after that period, such as POPPER, K. R., *Objective Knowledge. An Evolutionary Approach,* Clarendon Press, Oxford, 1972.

13 Cf. KUHN, TH. S., *The Structure of Scientific Revolutions*, The University of Chicago Press, Chicago, 1962 (2nd ed., 1970).

14 Cf. LAKATOS, I., *The Methodology of Scientific Research Programmes*, edited by John Worrall and Gregory Currie, Cambridge University Press, Cambridge, 1978.

15 Cf. LAUDAN, L., *Progress and its Problems. Towards a Theory of Scientific Growth*, University of California Press, Berkeley, 1977.

community requires for us to take into account social and political factors, and not only the internal constituents of science. Furthermore, the responsibility of the scientist goes beyond the endogenous aspects (honesty, originality, reliability, …) of scientific activity to reach the exogenous elements. Thus, the aims, processes and results are not mere individual ingredients but rather social factors.

1.2. "Naturalistic Turn," "Social Turn," and the Realist Conceptualizations

By connecting with the influential previous trends already pointed out (the "received view," Popper's approach, and the "historical turn") and, at the same time, opening new doors, the analysis of scientific progress was, in many ways, the central topic of discussion in general philosophy and methodology of science for a decade (1975-1985).[16] Certainly, the study of "scientific progress" was not restricted to the thinkers involved in the "historical turn." On this topic there are philosophic-methodological views of different kinds: naturalist, social, realist, etc. And, even though the moment of the turning point in favor of new trends is not so clear, it does seem clear that, from the beginning of the 1980's there emerges a bundle of new views on general philosophy and methodology of science that reach our present time.

For over two decades, the philosophic-methodological discussion has paid intense attention to several trends that belong to a "post-historical turn." Among them, there are three that are over-arching in the perspective on science: they conform what can be called the "naturalistic turn," the "social turn," and the realist conceptualizations. Nevertheless, there is some overlapping between two of them, because the discussion on scientific realism has frequently been interwoven in the debates with the other positions. At the same time, certain naturalisms are sociologically based, and this can lead to a view within the "social turn" (e.g., a relativistic option). In other words, it could be the case —and this happens *de facto*— that some authors can offer conceptions where there are elements of different kinds.

Nevertheless, each view —naturalist, social, or realist— has a different perspective on scientific progress. For Alexander Rosenberg, "progressivity" is a central feature of naturalism in philosophy of science. The base can be in history of science or sociology of science.[17] The social approach insists on the need for a characterization of "scientific progress" according to "external values" (social, cultural, economic, political, ecological, etc.). Thus, this position "affects philosophy of science as well as philosophy of technology. It includes a new vision of the aims, processes and results of scientific activities and technological doings, because the focus of attention is on several aspects of science and

---

[16] Cf. Harré, R. (ed.), *Problems of Scientific Revolution: Progress and Obstacles to Progress in the Sciences*, Oxford University Press, Oxford, 1975; Agazzi, E. (ed.), *Il concetto di progresso nella scienza*, Feltrinelli, Milan, 1976; Radnitzky, G. and Andersson, G. (eds.), *Progress and Rationality in Science*, Reidel, Dordrecht, 1978; Rescher, N., *Scientific Progress: A Philosophical Essay on the Economics of Research in Natural Science*, University of Pittsburgh Press, Pittsburgh, 1978; Dilworth, G., *Scientific Progress: A Study Concerning the Nature of the Relation between Successive Scientific Theories*, Reidel, Dordrecht, 1981; Shäfer, W. (ed.), *Finalization in Science: The Social Orientation of Scientific Progress*, Reidel, Dordrecht, 1983; Niiniluoto, I., *Is Science Progressive?*, Reidel, Dordrecht, 1984; and Pitt, J. (ed.), *Change and Progress in Modern Science*, Reidel, Dordrecht, 1985.

[17] Cf. Rosenberg, A., "A Field Guide to Recent Species of Naturalism," *The British Journal for the Philosophy of Science*, v. 47, (1996), p. 4.

technology that used to be considered as secondary, or even irrelevant. This *turn* highlights science and technology as *social undertakings* rather than intellectual contents."[18]

Meanwhile, any debate on scientific realism considers the necessity of "internal values" (linguistic, structural, epistemological, methodological, ontological, axiological, etc.), because it is a position where the constitutive elements of science are emphasized. Thus, when the realist discusses "scientific progress," he or she takes into account problems regarding objectivity in science and, commonly, the search for truth as well (or, at least, the possibility of truthlikeness). How this can be made varies according to the diverse orientations of realism, [19] because it is the case that there are a large variety of characterizations of "realism," in general, and of "scientific realism," in particular.[20]

2. Naturalistic Approaches

Historically, "naturalism" in philosophy appears as a cluster of positions rather than a single conception. Moreover, the "naturalistic turn" has involved a large number of naturalist approaches on science. There are several reasons for this phenomenon. On the one hand, it seems patent that "naturalism" is not a new view in philosophy when this "turn" starts.[21] Indeed, several versions of it were available within the general philosophical panorama.[22] And, on the other hand, the development of philosophic-methodological approaches on science has shown a possible version of naturalism according to each one of the constitutive elements of science (language, knowledge, method, activity, values, ...).

Hence, we can find several kinds of naturalism in science, among them: 1) semantic naturalism, where there is an acceptance of meaning as linguistic use, because meaning is based on a practice that can be described rather than prescribed; 2) epistemological naturalism, which accepts that human knowledge is well oriented and assumes a continuity between science and philosophy (and, then, that a metaphysical foundation of any of them is not needed); 3) methodological naturalism, where the progress of science (including the case of the social sciences) can be made through processes empirically tested according to criteria used in natural sciences; 4) ontological naturalism, which only accepts entities that in one way or another could be observable (i.e., it denies the legitimacy of unobservable entities such as "mind," "consciousness," and the like); and 5) axiological naturalism, where the scientific values are those that come from scientific practice.

However, according to the kind of support for these versions of naturalism (historical, psychological, biological, sociological, economic, etc.), there are additional varieties of

[18] Gonzalez, W. J., "The Relevance of Science, Technology and Society: The 'Social Turn'," in Gonzalez, W. J. (ed.), *Science, Technology and Society: A Philosophical Perspective*, Netbiblo, A Coruña, 2005, p. ix.

[19] It is also the case that a thinker has different versions of "realism," cf. Putnam, H., *The Many Faces of Realism*, Open Court, La Salle, IL, 1987; and Putnam, H., *Realism with a Human Face*, Harvard University Press, Cambridge, MA, 1990.

[20] Cf. Gonzalez, W. J., "El realismo y sus variedades: El debate actual sobre las bases filosóficas de la Ciencia," en Carreras, A. (ed.), *Conocimiento, Ciencia y Realidad,* Seminario Interdisciplinar de la Universidad de Zaragoza-Ediciones Mira, Zaragoza, 1993, pp. 11-58.

[21] A notorious case is Willard v. O. Quine's epistemological naturalism. Cf. Quine, W. v. O., "Epistemology Naturalized," in Quine, W. v. O., *Ontological Relativity and Other Essays,* Columbia University Press, New York, 1969, pp. 69-90.

[22] Cf. Strawson, P. F., *Skepticism and Naturalism: Some Varieties*, Columbia University Press, New York, 1985.

naturalistic approaches. This is especially clear in the cases of epistemological naturalism and methodological naturalism, because within them there can be a science (history, psychology, biology, sociology, etc.) that sustains the naturalist building up of scientific knowledge or scientific method. Thus, in the last two decades, we have seen naturalisms based on historical grounds (such as "normative naturalism"), psychological bases (such as several "cognitive approaches"), biological tenets (such as Darwinian characterizations of "adaptive rationality"), etc. Some of them are overlapped with the "social turn," whereas others are open to some dimensions of scientific realism.

2.1. Naturalism in the "Cognitive Turn"

An explicit defense of a "Philosophy of Science Naturalized" can be found in Ronald Giere in 1985.[23] His view is based on cognitive science, where the emphasis is on the psychology of the individual agents (perception, memory, language...) that develop science. Thus, a "cognitive theory of science" can be understood as a scientific explanation of science built upon the resources of cognitive science.[24] He offers a conception that does not dissolve the realist epistemological effort to represent the world: "There is room for a modest yet robust scientific realism that insists scientists are at least sometimes successful in their attempts to represent the causal structure of the world."[25]

These types of characterizations of philosophy and methodology of science based on cognitive psychology —in combination with other analysis, such as artificial intelligence and linguistic contributions— are sometimes called the "cognitive turn." In the case of Giere, it is clear that most of his approaches belong to the naturalistic realm.[26] But there are other cases where the cognitive emphasis moves towards a social epistemology.[27] Thus, the general framework of the cognitive approach belongs to the "social turn," insofar as the social aspects of knowledge have more weight than the individual characteristics of the agents that are doing science. Nevertheless, in both cases of "cognitive turn" —individual and social— a convergence with scientific realism is possible, insofar as an "inference to the best explanation" is accepted.

Among the main cognitive authors are Alvin Goldman[28] and Paul Thagard.[29] The latter is particularly relevant, because he tries to make the interest in the internal contents of science (conceptual elements of scientific theories) compatible with the external

23 Giere, R. N., "Philosophy of Science Naturalized," *Philosophy of Science,* v. 52, (1985), pp. 331-356.

24 Cf. Giere, R. N., *Explaining Science: A Cognitive Approach*, The University of Chicago Press, Chicago, 1988.

25 Giere, R. N., *Explaining Science: A Cognitive Approach*, p. xvi.

26 Cf. Giere, R. N. (ed.), *Cognitive Models of Science,* University of Minnesota Press, Minneapolis, 1992; and Giere, R. N., "The Cognitive Structure of Scientific Theories," *Philosophy of Science*, v. 61, (1994), pp. 276-296.

27 Cf. Fuller, S., *Social Epistemology,* Indiana University Press, Bloomington, 1988; and Fuller, S., *Philosophy, Rethoric, and the End of Knowledge: The Coming of Science and Technology Studies,* University of Wisconsin Press, Madison, 1993.

28 Cf. Goldman, A., *Epistemology and Cognition*, Harvard University Press, Cambridge, MA, 1986.

29 Cf. Thagard, P., *Computational Philosophy of Science,* The MIT Press, Cambridge, 1988; and Thagard, P., *Conceptual Revolutions,* Princeton University Press, Princeton, 1992. On this book, cf. Gonzalez, W. J., "P. Thagard: Conceptual Revolutions," *Mind,* v. 104, n. 416, (1995), pp. 916-919.

influence of the social environment.[30] Thus, his approach is not reduced to descriptive facets (psychological or sociological) of a purely naturalized perspective, because he also recognizes a prescriptive dimension related to scientific revolutions (rationality, coherence, etc.). Moreover, Thagard accepts objective elements in the interpretation of revolutions. This interest in objectivity distances him from the Kuhnian conversion, the primacy of the sociological explanation regarding the revolutionary change, and the linguistic emphasis to understand scientific revolutions in terms of translation.[31]

2.2. Normative Naturalism

Another kind of naturalism, besides the epistemological idea of continuity between science and philosophy, considers that both should use the same methodological principles of evaluation. Laudan has defended this "normative naturalism."[32] In his philosophic-methodological approach there are three different stages.[33] The *initial period* is "historiographical" (1965-1982), where the emphasis is on the methodological beliefs of a scientific community (a central point of *Progress and its Problems*).[34] Thus, he has a *transition* from the early metaintuitive criteria of decision-making of the "historical turn" to its explicit criticism (1983-1986).[35] The change is accompanied by the rejection of the demarcation —a central issue in Popper and Lakatos— as pseudoproblem.[36] This revision of his own thought is consolidated by the *posterior phase* of an explicit attempt to naturalize philosophy and methodology of science (from 1987 on), where the normativeness comes from scientist practice (mainly, historical).[37]

---

30 Cf. THAGARD, P., "Mind, Society and the Growth of Knowledge," *Philosophy of Science,* v. 61, (1994), pp. 629-645; and THAGARD, P., "Explaining Scientific Change: Integrating the Cognitive and the Social," in HULL, D., FORBES, M. and BURIAN, R. M. (eds.), *Proceedings of the 1994 Biennial Meeting of the Philosophy of Science Association,* Philosophy of Science Association, East Lansing, vol. 2, 1995, pp. 298-303.

31 Cf. GONZALEZ, W. J., "Towards a New Framework for Revolutions in Science," *Studies in History and Philosophy of Science*, v. 27, n. 4, (1996), p. 608.

32 Cf. LAUDAN, L., *Beyond Positivism and Relativism: Theory, Method, and Evidence,* Westview Press, Boulder, CO, 1996; LAUDAN, L., "Naturalismo normativo y el progreso de la Filosofía," in GONZALEZ, W. J. (ed), *El Pensamiento de L. Laudan. Relaciones entre Historia de la Ciencia y Filosofía de la Ciencia,* pp. 105-116; and LAUDAN, L., "Una Teoría de la evaluación comparativa de teorías científicas," in GONZALEZ, W. J. (ed), *El Pensamiento de L. Laudan. Relaciones entre Historia de la Ciencia y Filosofía de la Ciencia,* pp. 155-169.

33 A detailed analysis of these three stages is in GONZALEZ, W. J., "El giro en la Metodología de L. Laudan: Del criterio metaintuitivo al naturalismo abierto al relativismo débil," in VELASCO, A. (ed.) *Progreso, pluralismo y racionalidad. Homenaje a Larry Laudan*, Ediciones Universidad Nacional Autónoma de México, México D. F., 1999, pp. 105-131.

34 "The kind of things which count as empirical problems, the sorts of objections that are recognized as conceptual problems, the criteria of intelligibility, the standards for experimental control, the importance or weight assigned to problems, are all a function of the methodological-normative beliefs of a particular community of thinkers," LAUDAN, L., *Progress and its Problems*, p. 131.

35 Laudan recognizes openly this important change from the historiographical stage: "Since I no longer hold any brief for the meta-methodology of *Progress and its Problems*, I will refer to the author of those views in the third person," LAUDAN, L., "Some Problems Facing Intuitionistic Meta-Methodologies," *Synthese*, v. 67, (1986), p. 126.

36 Cf. LAUDAN, L., "The Demise of the Demarcation Problem," in COHEN, R. and LAUDAN, L. (eds.), *Physics, Philosophy and Psychoanalysis*, Reidel, Dordrecht, 1983, pp. 111-128; especially, p. 124.

37 Cf. GONZALEZ, W. J., "El naturalismo normativo como propuesta epistemológica y metodológica," in GONZALEZ, W. J. (ed.), *El Pensamiento de L. Laudan. Relaciones entre Historia de la Ciencia y Filosofía de la Ciencia,* pp. 5-57.

When Laudan adopts this normative naturalism the central tenets are openly non-realists, even though his epistemological pragmatism still accepts the idea of "objectivity," which conforms the bedrock for scientific realism. In addition, he criticizes relativism, but *de facto* his rejection affects one kind of it: "relativism stands or falls only on its stronger version."[38] Meanwhile, he does not dispute the thesis of a weak relativism. The difference is this: "weak relativism would be the position that there are some occasions when existing evidence does not permit a choice between certain rival theories. Strong relativism, by contrast, would hold that evidence is always powerless to choose between any pair of rivals."[39]

Several claims conform the main lines for articulation of this *normative naturalism*. a) From a structural point of view, Laudan establishes an identity between philosophical doings and scientific activity: "philosophy can, and should, make use of any of the forms of reasoning appropriate to scientific research."[40] b) Epistemologically, he rejects "truth" as scientific aim, because he considers it as quixotic: there are no epistemological criteria to determine whether we have reached it.[41] c) Methodologically, what is evaluated is dual: theory *cum* experience, and the key role for evaluation is in the evidence instead of being in the scientific beliefs.[42] d) Axiology of research is closely interrelated to methodology of science[43] (and both are inevitably intertwined in relations of mutual dependency with factual statements, within a reticulated model of scientific rationality).

Recently, Laudan has emphasized that he establishes a clear difference between "epistemic values" (understood as those related to truth and falsity) and cognitive values, i.e., "the role that issues of scope, generality, coherence, consilience, and explanatory power play in the evaluation of scientific theories."[44] In addition to the insistence on the need for cognitive values instead of epistemic values, he acknowledges the importance of social values, insofar as any human endeavor "is grounded in social processes of communication, negotiation, and consensus formation."[45]

Nicholas Rescher has offered a wider framework for the role of values in science than the previous approach. On the one hand, he defends a "holism of values in science,"[46] where the scientific values could be a matter of distinction ("internal" and "external,"

---

[38] LAUDAN, L., *Science and Relativism: Some Key Controversies in the Philosophy of Science*, The University of Chicago Press, Chicago, 1990, p. 56.

[39] LAUDAN, L., *Science and Relativism*, pp. 55-56.

[40] *Science and Relativism*, p. 99.

[41] "To tell scientists that they should seek true theories when (…) there is no way of certifying any theory as true, or 'truer than another', is to enjoin them to engage in a quixotic enterprise," LAUDAN, L., *Science and Relativism,* p. 53.

[42] On the role of evidence in Laudan's naturalism, cf. GONZALEZ, W. J., "El naturalismo normativo como propuesta epistemológica y metodológica," pp. 31-36.

[43] Cf. LAUDAN, L., "Progress or Rationality? The Prospect for Normative Naturalism," *American Philosophical Quarterly*, v. 24, n. 1, (1987), p. 29.

[44] LAUDAN, L., "The Epistemic, the Cognitive, and the Social," in MACHAMER, P. and WOLTERS, G. (eds.), *Science, Values, and Objectivity*, University of Pittsburgh Press, Pittsburgh, and Universitätsverlag, Konstanz, 2004, p. 20.

[45] LAUDAN, L., "The Epistemic, the Cognitive, and the Social," p. 22.

[46] Cf. *Personal communication*, 27 August 1998.

cognitive and social, ...) but this aspect does not authorize separations. On the other hand, he pays more attention to the existence of economic values in science than Laudan does. Moreover, this topic constitutes the *leit motiv* of two of Rescher's books: *Cognitive Economy*[47] *and Priceless Knowledge*?[48] This issue also appears in his reflection on rationality —in general, and scientific, in particular— from the point of view of the economic dimension.[49]

Even though Rescher is in tune with Laudan on some epistemological bases, such as a pragmatic vision of scientific knowledge, there are also differences between their viewpoints: first, because of the diverse emphasis on the question of limits that they put; and, second, due to the issue of the interconnection between cognitive goals of science and the rest of our goals as human beings. These aspects are highlighted by Rescher.[50] For him, one of the key values of science as a human cognitive project is "its selflimitation based in a need to recognize that there are limits to the extent to which this project can be realized."[51]

3. The Social Concern on Science

According to the new vision of the "social turn,"[52] the focus is on the social concern on science, i.e., scientific activity in a contextual setting. Thus, there are several important changes as to *what* should be studied —the objects of research—, *how* it should be studied —the method— and what the *consequences* for those studies are. The new focus of attention can be seen in many changes, and among them are several of special interest: a) from what science and technology are in themselves (mainly, epistemic contents) to how science and technology *are made* (largely, social constructions); b) from the language and structure of basic science to the characteristics of *applied science* and the *applications of science*; c) from technology as a feature through which human beings control their natural surroundings (a step beyond "technics" due to the contribution of science) to technology as a *social practice* and an *instrument of power*; and d) from the role of internal values necessary for "mature science" and "innovative technology" to the role of *contextual* or *external values* (cultural, political, economic …) of science and technology.

To be sure, this "social turn" is a move that covers a larger area and introduces a more radical scope than the preceding "historical turn," which was developed predominantly in the 60's and the 70's. On the one hand, the emphasis on *Science, Technology and Society*

---

47 Rescher, N., *Cognitive Economy: The Economic Perspectives of the Theory of Knowledge*, University of Pittsburgh Press, Pittsburgh, 1989.

48 Rescher, N., *Priceless Knowledge? Natural Science in Economic Perspective*, University Press of America, Savage, MD, 1996.

49 Cf. Rescher, N., "The Economic Dimension of Philosophical Inquiry," in Rescher, N., *Philosophical Reasoning. A Study in the Methodology of Philosophizing*, B. Blackwell, Oxford, 2001, Ch. 8, pp. 103-115. "Economic rationality is not the only sort of rationality there is, but it is an important aspect of overall rationality," Rescher, N., *Objectivity: The Obligations of Impersonal Reason*, Notre Dame University Press, Notre Dame, IN, 1997, p. 184.

50 Cf. Gonzalez, W. J., "Racionalidad científica y actividad humana. Ciencia y valores en la Filosofía de N. Rescher," in Rescher, N., *Razón y valores en la Era científico-tecnológica*, Paidós, Barcelona, 1999, pp. 21-22.

51 *Personal communication*, 27 August 1998.

52 This section follows Gonzalez, W. J., "The Relevance of Science, Technology and Society: The 'Social Turn'," pp. ix-x.

(*STS*) enlarges the domain in comparison with the contributions made by Kuhn, Lakatos, Laudan, etc. The role of historicity as a crucial element for the philosophical approach was analyzed mostly in the case of science. *De facto*, the major philosophers of that period paid little attention to technology. Furthermore, they customarily saw technology as an instrument that science uses for observation or experimentation. On the other hand, *STS* brings with it a more radical scope than the "historical turn," because that conception —including *The Structure of Scientific Revolutions*— still assumes that the internal contents of science have more weight than the external factors (social, cultural, political, economic, etc.).

In addition, there is a further enlargement introduced by the "social turn" in comparison with the "historical turn." *STS* considers the contributions of several disciplines, among them practical ethics, policy analysis, legal studies, sociology of science and sociology of technology, economics of science and economics of technology … Thus, the "social turn" includes more scientific contributions than history of science and history of technology. But the main interest is not in the intellectual history, either of science (e.g., of scientific theories) or of technology (e.g., on the changes in the *know-how*), but rather in contextual elements of the discoveries or improvements of science or technology (the search for fame, power relations, institutional traditions, etc.).

Science, Technology and Society or *Science and Technology Studies* are two ways of referring to an interdisciplinary endeavor.[53] *STS* combines the contributions of several disciplines and, accordingly, it uses different methodologies. Its object is not an isolated realm analyzed by a traditional kind of research, because it depends on views on science and technology developed in the last four decades. Indeed, *STS* has received increasing attention since the mid-1980's,[54] when the discussion explicitly included a third term: "technoscience." It is also a period where philosophy of technology increased progressively its presence in the realm of *STS*,[55] connecting technology with new areas for philosophical research (issues related to bioethics, environmental concerns, social problems, policy discussions, etc.).[56]

Since the constitution of *STS*, both philosophy of science and philosophy of technology have had a key role in this contemporary field. Their contributions are interconnected with contents of other disciplines. *De facto*, *STS* is a broad intellectual enterprise where several disciplines are involved: practical ethics, policy analysis, law, sociology, economics … The reason for this wide variety of contributions is clear: *STS* cannot be reduced to the theoretical study of science and technology, because it includes a *practical dimension* as

---

53 Cf. Gonzalez, W. J., "The Philosophical Approach to Science, Technology and Society," in Gonzalez, W. J. (ed.), *Science, Technology and Society: A Philosophical Perspective,* pp. 3-49; especially, pp. 3-4.

54 Some of the most influential views on *STS* had already started before the mid-1980's, cf. Barnes, B., *Scientific Knowledge and Sociological Theory,* Routledge and K. Paul, London, 1974; Latour, B. and Woolgar, S., *Laboratory Life: The Social Construction of Scientific Facts*, Princeton University Press, Princeton, NJ, 1979; Knorr-Cetina, K., *The Manufacture of Knowledge. An Essay on the Constructivist and Contextual Nature of Science*, Pergamon Press, Oxford, 1981; and Collins, H. M., *Frames of Meaning: The Sociological Construction of Extraordinary Science*, Routledge and K. Paul, London, 1982.

55 Cf. Ihde, D., "Has the Philosophy of Technology Arrived? A State-of-the-Art Review," *Philosophy of Science*, v. 71, n. 1, (2004), pp. 117-131.

56 Cf. Scharff, R. C. and Dusek, V. (eds.), *Philosophy and Technology: The Technological Condition*, Blackwell, Oxford, 2003.

well as a *social concern*. In Europe the first aspect is still dominant, whereas in the United States the second facet has a central relevance.[57]

Some philosophical views, such as certain versions of naturalism of sociological roots and most of the postmodern conceptions of science, have emphasized the sociological dimension of scientific activity. The existence in *STS* of three very influential positions within the sociology of science should be pointed out: a) the "Strong Program" of the Edinburgh school led by Barry Barnes[58] —now at the University of Exeter— and David Bloor,[59] based on some Kuhnian as well as Wittgensteinian ideas; b) the empirical program of relativism (EPOR) developed by Harry Collins,[60] which studies scientific controversies with an interpretative flexibility and analyzes their connections to the socio-cultural milieu; and c) the ethnomethodology —the study of the actors' network at the workplace— defended by the social constructivism of Bruno Latour[61] and Steve Woolgar.[62] All of them have gone far beyond the initial expectations of Thomas Kuhn, when he insisted on the historical character of science within a social setting of scientific communities, because he still gave more relevance to the internal constituents of science than to the external factors.[63]

4. The Interest in "Scientific Realism"

"Scientific realism" has been a central topic of discussion for several decades. This phenomenon can be explained as a combination of several factors: 1) the interest of a large set of philosophers in the development of different sorts of realism regarding science (semantic, epistemological, methodological, ontological, axiological, etc.); 2) the existence of another

57 In this regard, it is very clear in the contribution of Kristin Shrader-Frechette. Cf. Shrader-Frechette, K., "Objectivity and Professional Duties Regarding Science and Technology," in Gonzalez, W. J. (ed.), *Science, Technology and Society: A Philosophical Perspective*, pp. 51-79.

58 Cf. Barnes, B., *Interests and the Growth of Knowledge*, Routledge and K. Paul, London, 1977; Barnes, B., *T. S. Kuhn and Social Science,* Macmillan, London, 1982 (Columbia University Press, N. York, 1982); Barnes, B., *The Elements of Social Theory*, Princeton University Press, Princeton, 1995; and Barnes, B., Bloor, D. and Henry, J., *Scientific Knowledge. A Sociological Analysis*, The University of Chicago Press, Chicago, 1996.

59 Cf. Bloor, D., "Wittgenstein and Mannheim on the Sociology of Mathematics," *Studies in History and Philosophy of Science*, v. 4, (1973), pp. 173-191; Bloor, D., "Popper's Mystification of Objective Knowledge," *Science Studies*, v. 4, (1974), pp. 65-76; Bloor, D., *Knowledge and Social Imagery*, Routledge and K. Paul, London, 1976 (2nd ed., The University of Chicago Press, Chicago, 1991); Bloor, D., *Wittgenstein: A Social Theory of Knowledge*, Macmillan, London, 1983; and Bloor, D., *Wittgenstein, Rules and Institutions*, Routledge, London, 1997.

60 Cf. Collins, H. M., "An Empirical Relativist Programme in the Sociology of Scientific Knowledge," in Knorr-Cetina, K. D. and Mulkay, M. (eds.), *Science Observed: Perspectives in the Social Study of Science*, Sage, London, 1983, pp. 85-100; and Collins, H. M. and Pinch, T., *The Golem: What Everyone Should Know About Science*, Cambridge University Press, Cambridge, 1993.

61 Cf. Latour, B. and Woolgar, S., *Laboratory Life: The Social Construction of Scientific Facts*, 2nd ed., Princeton University Press, Princeton, NJ, 1986; Latour, B., *The Pasteurisation of France*, Harvard University Press, Cambridge, MA, 1988; and Latour, B., *We have Never been Modern,* Harvester, Brighton, 1993 (translated by C. Porter.)

62 Cf. Woolgar, S., "Critique and Criticism: Two Readings of Ethnomethodology," *Social Studies of Science*, v. 11, n. 4, (1981), pp. 504-514; Woolgar, S., *Science: The Very Idea,* Tavistock, London, 1988; and Woolgar, S. (ed.), *Knowledge and Reflexivity: New Frontiers in the Sociology of Knowledge*, Sage, London, 1988.

63 "I am among those who have found the claims of the strong program absurd: an example of deconstruction gone mad," Kuhn, Th. S., *The Road Since* Structure. *Philosophical Essays, 1970-1993, with an Autobiographical Interview*, edited by James Conant and John Haugeland, The University of Chicago Press, Chicago, 2000, p. 110.

important set of alternatives to realism in science (Michael Dummett's semantic anti-realism,[64] different versions of methodological instrumentalism —Laudan, the Structuralist Program, etc.—, and diverse versions of relativism); and 3) the acceptance in some cases of conceptions antagonist with realism, such as idealism, solipsism, and skepticism.[65]

Here the focus is on some central views on *scientific realism* as well as on some influential approaches based on realist grounds. Thus, following these lines, certain points of novelty and continuity in scientific realism will receive more attention in order to contribute to the picture of the contemporary framework on this topic. At the same time, it will be made clearer why a philosopher of science can be realist in some aspects (e.g., in ontology of science) while accepting an alternative view in other respects (e.g., in epistemology and in methodology of science). Meanwhile, the incompatibility of scientific realism and its antagonists (idealism, solipsism, and skepticism) seems obvious.

4.1. Central Views on Scientific Realism

There may be a number of realist approaches to science equivalent to the list of constitutive elements of science. In my judgment, the *characteristics of a science* are not simple, but they can be enumerated basically in some elements:[66] i) science possesses a specific language (with terms whose sense and reference are precise); ii) science is articulated in scientific theories with a well-patterned internal structure, which is open to later changes; iii) science is a qualified knowledge (with more rigor —in principle— than any other knowledge); iv) it consists of an activity that follows a method (normally deductive, although some authors accept the inductive method)[67] and it appears as a dynamic activity (of a self-corrective kind, which seeks to increase the level of truthlikeness).

Apart from these characteristics, there are others which have been emphasized in recent times: v) the reality of science comes from social action, and it is an activity whose nature is different from other activities in its assumptions, contents and limits; vi) science has aims —generally, cognitive ones— for guiding its research endeavor (in the formal sphere and in the empirical realm); and vii) it can have ethical evaluations insofar as science is a free human activity: values which are related to the process itself of research (honesty, originality, reliability …) or to its nexus with other activities of human life (social, cultural…).

Consequently, with this list of characteristics of a science, we have a large number of possibilities for scientific realism: semantic, logical, epistemological, methodological, ontological, axiological, and ethical. Commonly, the discussion pays more attention to some of these, mainly those versions related to language, knowledge, method, and reality. These factors involve a vision of the world as well as values (in the different realms: cognitive, ethical, social, economic, etc.) which influence scientific practice. In addition,

[64] Cf. Dummett, M., "Realism" (I), lecture at the *Oxford University Philosophical Society* on 8 March 1963. Reprinted in Dummett, M., *Truth and Other Enigmas*, Duckworth, London, 1978, pp. 145-165; and Dummett, M., "Realism" (II), *Synthese*, v. 52, (1982), pp. 55-112.

[65] A more detailed characterization of these three options can be found in Gonzalez, W. J., "El realismo y sus variedades: El debate actual sobre las bases filosóficas de la Ciencia," pp. 11-58; especially, pp. 11-13 and 23-30.

[66] Cf. Gonzalez, W. J., "The Philosophical Approach to Science, Technology and Society," pp. 10-11.

[67] Cf. Niiniluoto, I. and Tuomela, R., *Theoretical Concepts and Hypothetico-Inductive Inference*, Reidel, Dordrecht, 1973.

every possibility of approach to scientific realism (e.g., semantic, epistemological, or methodological) can receive different interpretations within a realist framework. Altogether —the list of levels and the number of interpretations in each— leads to a plethora of versions of scientific realism.

Given this highly complex sum of realist versions of science, it seems convenient to point out some central tenets of what can be assumed as "scientific realism:" 1) language in science can express or carry out an objective content, where meaning and reality —the reference— has an identifiable nexus; 2) scientific knowledge should be oriented towards truth, and so individual or social certainty is not good enough for a realist approach; 3) it could be scientific progress according to several criteria,[68] among them are those related to an improvement in the search for truth or in the levels of truthlikeness; 4) the world has, in principle, an independent existence but it is open to human knowledge, and its reality could be intelligible for the scientists through research; and 5) values in science (mainly, cognitive values) may have an objectivity and, thus, they are not merely reducible to a social construction.

Semantic realism in science can have different roots. Two of them have been influential: the theory of reference based on the insights of *Über Sinn und Bedeutung*, the famous conception of Gottlob Frege's philosophy of language (on sense and reference), updated and criticized by Dummett;[69] and the causal theory of reference developed by Hilary Putnam (and endorsed in one way or another by other philosophers, such as Saul Kripke or Keith Donnellan). Putnam's case is more relevant because he has articulated two versions: an initial approach on realism —before 1976—[70] and a posterior different conception —with influence of Dummett's anti-realism— called "internal realism."[71]

Epistemological realism also has a large diversity of options,[72] which vary according to the level of constructivism that they are ready to accept. The epistemological versions usually have methodological consequences.[73] Within the different possibilities on scientific realism from an epistemological point of view are those called "classic" (or naive realism), "materialist" (Hartry Field), "convergent" (Richard Boyd), "internal" (Hilary Putnam),

---

[68] A characterization of "scientific progress" is in GONZALEZ, W. J., "Progreso científico, Autonomía de la Ciencia y Realismo," *Arbor*, v. 135, n. 532, (1990), pp. 91-109; especially, pp. 96-100. On this notion, cf. NIINILUOTO, I., "Scientific Progress," *Synthese*, v. 45, (1980), pp. 427-464 (reprinted in NIINILUOTO, I., *Is Science Progressive*?, pp. 75-110); and NIINILUOTO, I., "Progress, Realism and Verisimilitude," in WEINGARTNER, P. and SCHURZ, G. (eds.), *Logic, Philosophy of Science and Epistemology*, Holder-Pichler-Tempsky, Vienna, 1987, pp. 151-161.

[69] Cf. DUMMETT, M., *Frege: Philosophy of Language,* Duckworth, London, 2nd ed. 1981 (1st ed. 1973); and DUMMETT, M., *The Interpretation of Frege's Philosophy,* Duckworth, London, 1981.

[70] Cf. PUTNAM, H., "Meaning and Reference," *Journal of Philosophy*, v. 70, (1973), pp. 699-711. Reprinted in SCHWARTZ, S. P. (ed.), *Naming, Necessity, and Natural Kinds*, Cornell University Press, Ithaca, 1977, pp. 118-132.

[71] Cf. PUTNAM, H., "Reference and Understanding," in PUTNAM, H., *Meaning and the Moral Sciences*, Routledge and K. Paul, London, 1978, pp. 97-119; and PUTNAM, H., "Reference and Truth," in PUTNAM, H., *Realism and Reason*, Cambridge University Press, Cambridge, 1983, pp. 69-86.

[72] They have competitors in the different versions of empiricism, either in a strict sense (such as Bas van Fraassen's "constructive empiricism") or in a broad sense (such as the variety of conceptions on Bayesianism).

[73] Commonly, a realist view on "scientific progress" concedes a particular relevance to a semantic dimension of science (and especially to the relations between truth and meaning). This is not the case in some naturalists perspectives, such as "normative naturalism." Since the "historical turn" the pragmatic interpretation of meaning has had a strong influence on many analysis of the problem of "progress" in science (mainly, in the "social turn").

and "critical scientific realism" (Ilkka Niiniluoto).[74] Some of these views are compatible with an epistemological naturalism, insofar as they assume the possibility of knowing the world as it is. But there are considerable discrepancies among realist and naturalist perspectives on some topics: the knowledge of unobservable things, the task of the critical attitude, the weight of intersubjective knowledge, etc.

Ilkka Niiniluoto is one of the most consistent realists. He endorses a *critical scientific realism*, which includes the following theses. R0 "At least part of reality is ontologically independent of human minds." R1 "Truth is a semantical relation between language and reality." R2 The concepts of truth and falsity are in principle applicable to all linguistic products of scientific enquiry, including observation reports, laws, and theories." R3 "Truth (together with some other epistemic utilities) is an essential aim of science." R4 "Truth is not easily accessible or recognizable, and even our best theories can fail to be true. Nevertheless, it is possible to approach the truth, and to make rational assessments of such cognitive progress." R5 "The best explanation for the practical success of science is the assumption that scientific theories in fact are approximately true or sufficiently close to the truth in the relevant aspects."[75]

Ontological realism is commonly a partner of epistemological realism. But this does not mean that an ontological realist is *eo ipso* an epistemological realist or vice versa. An ontological realist assumes the intelligility of the world. Thus, what is real (at least in the external world or realm independent from human mind) has many features to be discovered rather than to be constructed. For a realist, the structure and existence of real world is in itself autonomous from the act of thought of human mind. At least in the case of natural world, the ontological status is not an invention or the result of a social construction: that world has the constitutive character of otherness. The ontological object is not the epistemological subject.

Ian Hacking has insisted on the limitations of the social constructivist approach. His main target has been on what can be called "socially constructed" and on what is the task of science regarding that area. [76] It seems clear that he does not adopt a classic metaphysical realism or a naive scientific realism. Hacking's conception is a kind of critical realism: he does not accept an essence is to be discovered, and he also criticizes the possibility of a description to reflect an inherent internal structure of the things. But he admits the knowability and intelligibility of processes and objects (i.e., that reality other than the researchers'). Thus, regarding the quarks in the atoms, he maintains their ontological character: "quarks, the objects themselves, are not constructs, are not social, and are not historical."[77]

4.2. Some Influential Approaches Based on Realist Grounds

Particularly interesting are the cases of some influential philosophers of science, such as John Worrall and Philip Kitcher, who endorse scientific realist positions after their

[74] On these views, cf. GONZALEZ, W. J., "El realismo y sus variedades: El debate actual sobre las bases filosóficas de la Ciencia," pp. 11-58; especially, pp. 31-32 and 37-50.

[75] Cf. NIINILUOTO, I., *Critical Scientific Realism*, Clarendon Press, Oxford, 1999, p. 10.

[76] Cf. HACKING, I., *The Social Construction of* What?, Harvard University Press, Cambridge, MA, 1999.

[77] HACKING, I., *The Social Construction of* What?, p. 30.

involvement in the "historical turn" —the first author— and in the "naturalist turn" (the second thinker). They have made explicit versions of scientific realism in a personal way, commonly in discussion with other well-known specialists. Moreover, they have shown a transition to new topics in their careers (such as philosophy of medicine or recent problems related to philosophy of biology with a clear social repercussion).

John Worrall started his intellectual trajectory working within the "historical turn" due to his close relation with Lakatos, whose methodology of programs he considers superior to Popperian falsificationism.[78] Later on, he has developed a "structural realist" view,[79] and, at the same time, he has remained committed to the study of real events of the history of science. Thus, Worrall has very closely followed historical cases of the role of successful predictions in the acceptance of scientific theories.[80] These analyses are sometimes used to articulate a view of scientific change which is critical with an instrumentalist view of scientific theories.[81] In addition, he has paid attention to objective standards of rationality, criticizing some authors of the "social turn," such as David Bloor and Barry Barnes.[82]

Worrall has defended scientific realist views in different ways, mostly in the epistemological and methodological domains. a) He has argued consistently for a view of the development of science as a predominantly rational affair.[83] This includes a criticism of Kuhnian "scientific revolutions" and the search for a characterization of scientific progress where the role of logic could be relevant.[84] b) He has tried to fill the gap of Lakatos's "positive heuristics," looking for historical cases of procedures of scientific discoveries that can be reconstructed logically.[85] c) Worrall has criticized Laudan's methodological conceptions in several key points: i) in favor of the relevance of the considerations of theory-discovery;[86] and ii) against the intrinsic variability of the scientific methodology.[87]

---

[78] Cf. WORRALL, J., "The Ways in Which the Methodology of Scientific Research Programmes Improves on Popper's Methodology," in RADNITZKY, G. and ANDERSSON, G. (eds.), *Progress and Rationality in Science,* pp. 45-70.

[79] Cf. WORRALL, J., "Structural Realism: The Best of Both Worlds?," *Dialectica,* v. 43, n. 1-2, (1989), pp. 99-124.

[80] Cf. WORRALL, J., "Fresnel, Poisson and the White Spot: The Role of Successful Predictions in the Acceptance of Scientific Theories," in GOODING, D., PINCH, T. and SCHAFFER, S. (eds.), *The Uses of Experiment: Studies of Experimentation in Natural Science*, Cambridge University Press, Cambridge, 1989, pp. 135-157; and SCERRI, E. R. and WORRALL, J., "Prediction and the Periodic Table," *Studies in History and Philosophy of Science*, v. 32, n. 2, (2001), pp. 407-452.

[81] Cf. WORRALL, J., "Scientific Realism and Scientific Change," *Philosophical Quarterly,* v. 32, (1982), pp. 201-231.

[82] Cf. WORRALL, J., "Rationality, Sociology and the Symmetry Thesis," *International Studies in the Philosophy of Science*, v. 4, n. 3, (1990), pp. 305-319.

[83] Cf. WORRALL, J., "Scientific Revolutions and Scientific Rationality: The Case of the 'Elderly Holdout'," in SAVAGE, C. WADE (ed.), *Scientific Theories,* University of Minnesota Press, Minneapolis, 1990, pp. 319-354.

[84] Cf. WORRALL, J., "Realismo, racionalidad y revoluciones," *Agora,* v. 17, n. 2, (1998), pp. 7-24.

[85] Cf. WORRALL, J., "Programas de investigación y heurística positiva: Avance respecto de Lakatos," in GONZALEZ, W. J. (ed.), *La Filosofía de Imre Lakatos: Evaluación de sus propuestas,* pp. 247-268.

[86] Cf. WORRALL, J., "Scientific Discovery and Theory-Confirmation," in PITT, J. C. (ed.), *Change and Progress in Modern Science,* pp. 301-331.

[87] Cf. WORRALL, J., "The Value of a Fixed Methodology," *The British Journal for the Philosophy of Science,* v. 39, (1988), pp. 263-275; and WORRALL, J., "Fix It and Be Damned: A Reply to Laudan," *The British Journal for the Philosophy of Science,* v. 40, (1989), pp. 376-388.

Even though there is a large number of conceptions with special emphasis on scientific realism, it is also possible that a philosopher of science can defend a version of realism (semantic, epistemological, methodological, or ontological) and, at the same time, adopt some naturalistic claims or make very clear statements about the social concern on science. This is the case of Philip Kitcher, who has maintained a realistic conception of scientific progress,[88] and later on has developed a systematic approach to deal with the problems raised by science within a democratic society.[89]

Kitcher started his academic career with philosophy of mathematics. He made an important reflection on the nature of mathematical knowledge,[90] which was based on a naturalist attitude of empiricist roots.[91] In this initial stage, he accepts a naturalism in human knowledge: in the act of knowing, we should have the true belief that the knowledge is carried on by us in a correct way.[92] Furthermore, he adds a social component to the epistemology: "This theory (…) adscribes to present mathematical community and to previous communities an epistemological significance with which they are not usually credited."[93]

Later on, Kitcher recognizes the return of naturalism to the front line of the philosophic-methodological discussion after the intense period of the "historical turn."[94] But his acceptance of naturalist claims is made commonly in accordance with central tenets of scientific realism, either epistemological or methodological. This commitment to realism is clear in his main book on scientific progress: *The Advancement of Science*.[95] Thereafter, Kitcher moves softly to a modest realist position, when his concern to social dimension of science grows. Moreover, in recent years, his views are increasingly a blend of realism and pragmatism, at least in the issues on truth related to epistemology and methodology of science.

Science, Truth, and Democracy lies within this recent position, insofar as Kitcher connects truth and success: "defending the strategy of inferring truth from success against a bevy of historical examples is only the first step in turning back empiricist challenges to realism."[96] In addition, there is a commitment to ontological realism when he makes a "minimal realist claim:" "The sciences sometimes deliver the truth about a world independent of human cognition, and they inform us about constituents of that world that are remote from human observation."[97]

---

[88] Cf. KITCHER, PH., *The Advancement of Science: Science without Legend, Objectivity without Illusions,* Oxford University Press, N. York, 1993.

[89] That question is the focus of the second part of his book *Science, Truth, and Democracy*. In the first part, Kitcher includes a modest version of scientific realism.

[90] Cf. KITCHER, PH., *The Nature of Mathematical Knowledge*, Oxford University Press, Oxford, 1983.

[91] "I hope to show that Mill's views about arithmetic can be developed into a satisfactory theory of arithmetical truth and arithmetical knowledge," KITCHER, PH., "Arithmetic for the Millian," *Philosophical Studies*, v. 37, (1980), p. 215.

[92] Cf. GONZALEZ, W. J., "'Verdad' y 'prueba' ante el problema del progreso matemático," in MARTINEZ FREIRE, P. (ed.), *Filosofía Actual de la Ciencia,* Publicaciones Universidad de Málaga, Málaga, 1998, pp. 307-346; especially, p. 309.

[93] KITCHER, PH., *The Nature of Mathematical Knowledge*, p. 7.

[94] Cf. KITCHER, PH., "The Naturalist Returns," *Philosophical Review*, v. 101, (1992), pp. 53-114.

[95] Cf. *The Advancement of Science*, Chapter 5, pp. 127-177.

[96] KITCHER, PH., *Science, Truth, and Democracy,* p. 19.

[97] *Science, Truth, and Democracy,* p. 28.

5. The Increased Emphasis on Views Based on Theories of Probability: The Case of Bayesianism

Besides the "turns" already pointed out (naturalistic, social, ...), there is an increased emphasis on views based on theories of probability.[98] This happens at the same time as those options (and, sometimes, they are developed in the same university department). Bayesianism has been particularly influential among these theories of probability. Certainly, it is not a new conception of probability.[99] The revitalization of this view is connected to important books, such as *Scientific Reasoning: The Bayesian Approach*, published by Colin Howson and Peter Urbach in 1989.[100] Thereafter, new insights on epistemology and methodology of Bayesianism appeared, including relevant critical examinations.[101]

Epistemologically, the Bayesian approach of Howson and Urbach includes a subjective interpretation of the probability calculus and acceptance of the legitimacy of a mathematical expression of degree of belief. The objective probability is seriously criticized because, in addition to the problem of arbitrary assumptions, there is no systematic account of how their probabilistic predictions are to be empirically evaluated. The Theorem of Bayes is understood as the means to cope with the problems on prediction: to avoid arbitrary assumptions and to make the predictions valuable from an empirical point of view. This Bayesian approach involves a rejection of Popper's central tenets. In addition, it leads towards a position that, to some extent, can be located between Carl G. Hempel's logical empiricism and Kuhnian views.[102]

Methodologically, the Bayesian approach —understood in the personalist way— includes three elements. i) The role of the subject in the research process is emphasized, because the method of investigating has its starting point in intuitive principles of the scientist, which has a direct expression in the introduction of prior probabilities. ii) Induction is not only something that is accepted for developing science, but also it is explicitly defended against the critics of inductive probability: "All the alternative accounts of inductive inference —like Popper's or Fisher's— achieve their explanatory goals, where they achieve them at all, only at the cost of quite arbitrary stipulations."[103] iii) Deterministic events are not the "ideal" cases for the explanation of scientific

[98] An analysis of the main views on probability theories can be found in Gillies, D. A., *Philosophical Theories of Probability*, Routledge, London, 2000 (reprinted in 2003).

[99] It is based on Bayes, Th., "An Essay Towards Solving a Problem in the Doctrine of Chances," *Philosophical Transactions of the Royal Society*, v. 53, (1763), pp. 370-418. Richard Price —fellow of the *Royal Society*— submitted it for publication two years after Thomas Bayes's death.

[100] Howson, C. and Urbach, P., *Scientific Reasoning: The Bayesian Approach*, Open Court, La Salle, IL, 1989 (reprinted in 1990). There is a second edition (Open Court, La Salle, IL, 1993) and a third one (2006). (The references here will be from the original edition). This book was analyzed in Gonzalez, W. J., "El razonamiento científico desde una perspectiva bayesiana," *LLull*, v. 15, (1992), n. 28, pp. 209-218.

[101] Cf. Earman, J., *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*, The MIT Press, Cambridge, MA, 1992.

[102] Cf. Salmon, W. C., "Rationality and Objectivity in Science, or Tom Kuhn Meets Tom Bayes," in Savage, C. Wade (ed.), *Scientific Theories,* pp. 175-204.

[103] Howson, C. and Urbach, P., *Scientific Reasoning: The Bayesian Approach*, p. 257. From the beginning of the book —in the first edition—, there is an explicit criticism of approaches developed by Karl Popper and Ronald A. Fischer in favor of objectivity and classical methods of statistical inference. Cf. *Scientific Reasoning: The Bayesian Approach*, p. 11.

phenomena; they could be in a second level, being reinterpreted by theory of probability, because the probabilistic approach would be more embracing than the methodological perspective inspired in deterministic scientific theories.

This Bayesian approach links prediction with confirmation; moreover, prediction enhances confirmation. At the same time, this position is built up with a throughout probabilistic perspective, because "random phenomena defy any attempt at categorical prediction."[104] According to the Bayesian view, scientific inference is the evaluation of hypotheses in the light of empirical evidence. Predictions are one element among others of such activity; they belong to theories and appear in hypotheses. On the one hand, predicting is a way of testing and confirming theories; and, on the other hand, predictions are consequences of a logical process in the hypothesis: they are consequences that are drawn from the hypothesis (in order to evaluate it by observational or experimental evidence.) So, there are at least two ingredients: an adequate inference, following the probability calculus, and evaluation of the prediction by evidence that can confirm or disconfirm the hypothesis (and, eventually, the theory). This approach affects deterministic theories as well as statistical ones; in both cases there are evaluations of claims —some of them predictive— by calculating the probability of the claim in the light of given information. By means of Bayes's Theorem we know why and under what circumstances a hypothesis is confirmed by its consequences.

Within this general scope, some features of prediction are important for the Bayesian perspective. 1) The main goal in science is not to make predictions, because —contrary to Lakatos's position— there is no rational basis for giving more relevance to predictions than to other aspects in science: "a theory that produces a false prediction may still remain very probable."[105] 2) Predictions have a specific role in the "corroboration" of theories, but this cannot be understood in Popper's sense, because if the predictions fail, logic does not compel one to infer that the main theory is false. In fact, scientists consider evidence in a different way: a) many deterministic theories have no directly checkable deductive consequences and the predictions by which they are tested are necessarily drawn up with only the assistance of auxiliary theories; b) some theories are explicitly probabilistic and so make no categorical predictions; c) in some deterministic theories, instead of checking logical consequences, there is an examination of experimental effects which are predicted only with certain probability. 3) There are no especial reasons for preferring prediction to accommodation, so the problem that Hempel has raised is solved by means of deciding in each case which is pertinent.[106] "The plausibility of the thesis that predictions always glean more support than accommodations rests … on nothing more than invalidly generalising from this special case in which the thesis is true."[107] 4) The Popperian distinction between "prediction" and "prophecy" is avoidable, insofar as "prophecy" can be understood as something which is predicted: it corresponds to some degree of belief that can receive a mathematical expression according to the probability calculus. This could be one corollary

[104] Howson, C. and Urbach, P., *Scientific Reasoning: The Bayesian Approach*, p. 220.

[105] *Scientific Reasoning: The Bayesian Approach*, p. 97.

[106] Cf. Hempel, C. G., *Philosophy of Natural Science*, Prentice Hall, Englewood Cliffs, NJ, 1966, Chapter 4, pp. 33-46.

[107] Howson, C. and Urbach, P., *Scientific Reasoning: The Bayesian Approach*, p. 282.

more of Dutch Book Argument that stresses the thesis that strength of belief can be measured numerically and that such a measure satisfies the axioms of the probability calculus.

Following these general features, it seems that the Bayesian approach tries to avoid controversies such as "logic of discovery" versus "psychology of invention" or symmetry versus asymmetry between explanation and prediction. On the one hand, the "psychological tendencies are so far from being useless for scientific purposes that they form the everyday logic —and a genuine logic at that— of scientific inference."[108] On the other hand, in spite of the usual trend, there is no systematic treatment of topics like "explanation" and its relation with "prediction" in the original version of *Scientific Reasoning: The Bayesian Approach*.

For the Bayesian view —against Popper's and Fisher's criticisms—, logical and psychological aspects appear together —in some way or another— in scientific prediction; and there is no elbow room for distinguishing —in any relevant sense— symmetries and asymmetries between "explanation" and "prediction": both are logical consequences to be evaluated empirically. Nevertheless, an important feature of inductive reasoning is that "the scientist often prefers a theory which explains the data imperfectly (…) to an alternative (…) which predicts them with complete accuracy."[109]

In the Bayesian approach prediction is an ingredient of the method of investigating: from a hypothesis (deterministic or probabilistic) some logical consequences are drawn, which must be checked as true or not; if the empirical evaluation is in agreement with what has been predicted, the hypothesis is confirmed. It is not necessary to test a hypothesis indefinitely, although it is not an easy matter to predict the precise point beyond which further predictions of the hypothesis are sufficiently probable as not to be worth examining. "Repetitions of an experiment often confirm a general theory only to a limited extent, for the predictions verified by means of a given kind of experiment (that is, an experiment designed to a specified pattern) do normally follow from and confirm a much restricted version of the predicting theory."[110]

6. The Structure and Origin of the Present Volume

Each of the main philosophical and methodological trends of recent decades (the naturalist approaches, the views based on social concern, the conceptions with special emphasis on scientific realism, or the Bayesian perspective) can be found in this book. In some cases the analysis remains as a general stance, whereas in a large number of cases the attention shifts to specific sciences (medicine, biology, neuroscience, economics, psychology, mathematics...). In this feature some *elements of novelty* can be seen in comparison with the previous tendencies: the emphasis on problems of concrete subjects rather than on offering over-arching views (like "paradigms," "research programs," "research traditions," etc.), and the interest in focusing subjects other than physics (which used to have an special attraction to logical empiricists, critical rationalists, supporters of scientific research programs, etc.).

[108] Howson, C. and Urbach, P., *Ibidem*, p. 56.

[109] *Scientific Reasoning: The Bayesian Approach*, p. 118.

[110] Howson, C. and Urbach, P., *Scientific Reasoning: The Bayesian Approach*, p. 83.

Basically, the structure of the book is summarized in its three parts: I) Methodological Approaches on Central Problems of Science; II) Epistemological Issues related to a General Framework; and III) Focal Philosophical Problems in Empirical and Formal Sciences. Following the scheme, which analyzes first the broad topics in order to move later on to the more specific issues, every part has two sections. Thus, in the case of part I, these are "Scientific Reasoning and Discovery in Science," and "Scientific Testing;" in part II: "The Examination of Determinism and the Analysis of Life," and "Social Epistemology and the Cognitive Relation Science-Technology;" and in part III: "Neuroscience and Psychology," and "Mathematical Activity and Philosophical Problems."

Originally, these questions were studied through papers presented in a summer course of the International University Menéndez Pelayo (4-8 July 2005), which was at the same time a graduate course of the doctoral program of the University of Valencia related to "Language, Reason, and History." This program was awarded special recognition ("Quality mention") by the Spanish Ministry of Education and Science. The title of the course was the same as this book ("Perspectivas actuales en Filosofía y Metodología de la Ciencia").

Among the speakers on the course were the authors of the papers: Colin Howson (*London School of Economics*), Donald Gillies (*University College London*), John Worrall (*London School of Economics*), Wenceslao J. González (*University of A Coruña*), Frank Wuketits (*University of Vienna*), Emilio Muñoz (*Higher Council of Scientific Research*), Anna Estany (*Autonomous University of Barcelona*), Peter Machamer (*University of Pittsburgh*), José Sanmartín (*University of Valencia*), Javier de Lorenzo (*University of Valladolid*), and Jesús Alcolea (*University of Valencia*). In addition, this book includes a paper from Peter Clark (*University of St. Andrews*), who was in the original list of participants. Now he offers us his relevant contribution to the contents of this volume.

All the above presented the papers that have been adapted for the book. Thus, all papers have been reviewed for publication. This aspect has contributed to displaying a more systematic exposition for the reader. In addition to the lectures, there were two round tables during the course, with the idea of offering a broader picture to the audience. The first was on "Philosophy of Life Sciences." The speakers were Donald Gillies, Franz Wuketits, José Hernández Yago and Andrés Moya. Meanwhile the second round table was on "Philosophy of Sciences of the Artificial." In this case Anna Estany, Emilio Muñoz, José Hernández Orallo, and Peter K. Machamer discussed the topic.

After the contextualization of this volume —the topics and the origin of the book—, it is time to give thanks to those who have made it possible. Among them are the people and institutions involved in the course where most of these papers were presented. Thus, I wish to express my gratitude towards the International University Menéndez Pelayo, especially its director in Valencia (Vicente Bellver), and the doctoral program of the University of Valencia on "Language, Reason, and History," mainly to its codirector (Rafael Beneyto). My acknowledgement goes also to the Queen Sofia Center for the Study of Violence, for its support to the course, particularly its director: José Sanmartín.

I wish to express my warm gratitude to each one of the speakers on the course (Colin Howson, Donald Gillies, John Worrall, Frank Wuketits, Emilio Muñoz, Anna Estany, Peter Machamer, José Sanmartín, Javier de Lorenzo, and Jesús Alcolea) for accepting the

invitation to the event and for their papers. My special recognition goes to John Worrall, due to the efforts made to be in Valencia those days, and to Peter Clark, who was unable to be there that week, but who has sent his paper for publication. In addition, I would like to thank the speakers of the round tables: José Hernández Yago (*Polytechnic University of Valencia*), Andrés Moya (*University of Valencia*), and José Hernández Orallo (*Politechnic University of Valencia*).

Finally, I want to make explicit my recognition of Jesús Alcolea. He was a key figure of this activity in Valencia, due to his being a member of the Department of Logic and Philosophy of Science, which offers the doctoral program on "Language, Reason, and History." Furthermore, he was delighted with the idea of the course, and his assistance in organizational matters was decisive in the success of this event of the International University Menéndez Pelayo and the University of Valencia. I would like also to thank him for his contribution to the editing of this volume.

7. Bibliography

Because of the wealth of details involved in the analysis of the novelty and continuity in philosophy and methodology of science, the present bibliography is oriented in two directions: i) to offer a systematic account of what has been pointed out in the footnotes, with a limited range of additions; and ii) to make explicit some places where the bibliographical information can be enlarged to complete the panorama described in the text of this paper.

Concerning the second aspect, the references will habitually be to publications of the last ten years, thus seeking updated versions of the topics. Due to the obvious connections of the trends analyzed in this paper with the issues studied in the collection *Gallaecia. Studies in Contemporary Philosophy and Methodology of Science,* there are explicit mentions here to the bibliographical sections of some of those monographs that I have edited.

These sections are in the following books: *El Pensamiento de L. Laudan. Relaciones entre Historia de la Ciencia y Filosofía de la Ciencia* (1998), pp. 46-57; *La Filosofía de Imre Lakatos: Evaluación de sus propuestas* (2001), pp. 87-103; *Diversidad de la explicación científica* (2002), pp. 36-49; *Análisis de Thomas Kuhn: Las revoluciones científicas* (2004), pp. 80-103; *Karl Popper: Revisión de su legado* (2004), pp. 108-194; and *Science, Technology and Society: A Philosophical Perspective* (2005), pp. 37-49. In addition to the collection *Gallaecia*, it should be pointed out the bibliographical section of another volume that I have edited: *Racionalidad, historicidad y predicción en Herbert A. Simon* (2003), pp. 28-63. Finally, the book published by Nicholas Rescher under the title *Razón y valores en la Era científico-tecnológica* (1999), which I edited, contains a large amount of his publications as well as a bibliography about his thought (pp. 35-44).

Agazzi, E. (ed.), *Il concetto di progresso nella scienza*, Feltrinelli, Milan, 1976.

Balzer, W., Moulines, C. U. and Sneed, J. D., *An Architectonic for Science. The Structuralist Program*, Reidel, Dordrecht, 1987.

Barnes, B., *Scientific Knowledge and Sociological Theory*, Routledge and K. Paul, London, 1974.

Barnes, B., *Interests and the Growth of Knowledge*, Routledge and K. Paul, London, 1977.

BARNES, B., *T. S. Kuhn and Social Science,* Macmillan, London, 1982 (Columbia University Press, New York, 1982).

BARNES, B., *The Elements of Social Theory*, Princeton University Press, Princeton, 1995.

BARNES, B., BLOOR, D. and HENRY, J., *Scientific Knowledge. A Sociological Analysis*, The University of Chicago Press, Chicago, 1996.

BAYES, TH., "An Essay Towards Solving a Problem in the Doctrine of Chances," *Philosophical Transactions of the Royal Society*, v. 53, (1763), pp. 370-418.

BLOOR, D., "Wittgenstein and Mannheim on the Sociology of Mathematics," *Studies in History and Philosophy of Science*, v. 4, (1973), pp. 173-191.

BLOOR, D., "Popper's Mystification of Objective Knowledge," *Science Studies*, v. 4, (1974), pp. 65-76.

BLOOR, D., *Knowledge and Social Imagery*, Routledge and K. Paul, London, 1976 (2nd ed., The University of Chicago Press, Chicago, 1991).

BLOOR, D., *Wittgenstein: A Social Theory of Knowledge*, Macmillan, London, 1983.

BLOOR, D., *Wittgenstein, Rules and Institutions*, Routledge, London, 1997.

CLARK, P. and HAWLEY, K. (eds.), *Philosophy of Science Today*, Oxford University Press, Oxford, 2003.

COLLINS, H. M., *Frames of Meaning: The Sociological Construction of Extraordinary Science*, Routledge and K. Paul, London, 1982.

COLLINS, H. M., "An Empirical Relativist Programme in the Sociology of Scientific Knowledge," in KNORR-CETINA, K. D. and MULKAY, M. (eds.), *Science Observed: Perspectives in the Social Study of Science*, Sage, London, 1983, pp. 85-100.

COLLINS, H. M. and PINCH, T., *The Golem: What Everyone Should Know About Science*, Cambridge University Press, Cambridge, 1993.

DILWORTH, G., *Scientific Progress: A Study Concerning the Nature of the Relation between Successive Scientific Theories*, Reidel, Dordrecht, 1981.

DUMMETT, M., "Realism" (I), lecture at the *Oxford University Philosophical Society* on 8 March 1963. Reprinted in DUMMETT, M., *Truth and Other Enigmas*, Duckworth, London, 1978, pp. 145-165.

DUMMETT, M., *Frege: Philosophy of Language*, Duckworth, London, 2nd ed. 1981 (1st ed. 1973).

DUMMETT, M., *The Interpretation of Frege's Philosophy*, Duckworth, London, 1981.

DUMMETT, M., "Realism" (II), *Synthese*, v. 52, (1982), pp. 55-112.

EARMAN, J., *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*, The MIT Press, Cambridge, MA, 1992.

FULLER, S., *Social Epistemology*, Indiana University Press, Bloomington, IN, 1988.

FULLER, S., *Philosophy, Rethoric, and the End of Knowledge: The Coming of Science and Technology Studies*, University of Wisconsin Press, Madison, 1993.

GIERE, R. N., "Philosophy of Science Naturalized," *Philosophy of Science*, v. 52, (1985), pp. 331-356.

GIERE, R. N., *Explaining Science: A Cognitive Approach*, The University of Chicago Press, Chicago, 1988.

GIERE, R. N. (ed.), *Cognitive Models of Science*, University of Minnesota Press, Minneapolis, 1992.

GIERE, R. N., "The Cognitive Structure of Scientific Theories," *Philosophy of Science*, v. 61, (1994), pp. 276-296.

GILLIES, D. A., *Philosophy of Science in the Twentieth Century. Four Central Themes*, B. Blackwell, Oxford, 1993.

GILLIES, D. A., *Philosophical Theories of Probability*, Routledge, London, 2000 (reprinted in 2003).

GOLDMAN, A., *Epistemology and Cognition*, Harvard University Press, Cambridge, MA, 1986.

GONZALEZ, W. J., "Progreso científico, Autonomía de la Ciencia y Realismo," *Arbor*, v. 135, n. 532, (1990), pp. 91-109.

GONZALEZ, W. J., "El razonamiento científico desde una perspectiva bayesiana," *LLull*, v. 15, (1992), n. 28, pp. 209-218.

GONZALEZ, W. J., "El realismo y sus variedades: El debate actual sobre las bases filosóficas de la Ciencia," in CARRERAS, A. (ed.), *Conocimiento, Ciencia y Realidad*, Seminario Interdisciplinar de la Universidad de Zaragoza-Ediciones Mira, Zaragoza, 1993, pp. 11-58.

GONZALEZ, W. J., "P. Thagard: Conceptual Revolutions," *Mind*, v. 104, n. 416, (1995), pp. 916-919.

GONZALEZ, W. J., "Reichenbach's Concept of Prediction," *International Studies in the Philosophy of Science*, v. 9, n. 1, (1995), pp. 37-58.

GONZALEZ, W. J., "Towards a New Framework for Revolutions in Science," *Studies in History and Philosophy of Science*, v. 27, n. 4, (1996), pp. 607-625.

GONZALEZ, W. J. (ed.), *El Pensamiento de L. Laudan. Relaciones entre Historia de la Ciencia y Filosofía de la Ciencia*, Publicaciones Universidad de A Coruña, A Coruña, 1998.

GONZALEZ, W. J., "El naturalismo normativo como propuesta epistemológica y metodológica," in GONZALEZ, W. J. (ed.), *El Pensamiento de L. Laudan. Relaciones entre Historia de la Ciencia y Filosofía de la Ciencia*, Publicaciones Universidad de A Coruña, A Coruña, 1998, pp. 5-57.

GONZALEZ, W. J., "'Verdad' y 'prueba' ante el problema del progreso matemático," in MARTINEZ FREIRE, P. (ed.), *Filosofía Actual de la Ciencia*, Publicaciones Universidad de Málaga, Málaga, 1998, pp. 307-346.

GONZALEZ, W. J., "El giro en la Metodología de L. Laudan: Del criterio metaintuitivo al naturalismo abierto al relativismo débil," in VELASCO, A. (ed.) *Progreso, pluralismo y racionalidad. Homenaje a Larry Laudan*, Ediciones Universidad Nacional Autónoma de México, México D. F., 1999, pp. 105-131.

GONZALEZ, W. J., "Racionalidad científica y actividad humana. Ciencia y valores en la Filosofía de N. Rescher," in RESCHER, N., *Razón y valores en la Era científico-tecnológica*, Paidós, Barcelona, 1999, pp. 11-44.

GONZALEZ, W. J. (ed.), *La Filosofía de Imre Lakatos: Evaluación de sus propuestas*, UNED, Madrid, 2001.

GONZALEZ, W. J. (ed.), *Diversidad de la explicación científica*, Ariel, Barcelona, 2002.

GONZALEZ, W. J. (ed.), *Racionalidad, historicidad y predicción en Herbert A. Simon*, Netbiblo, A Coruña, 2003.

GONZALEZ, W. J. (ed.), *Análisis de Thomas Kuhn: Las revoluciones científicas*, Trotta, Madrid, 2004.

GONZALEZ, W. J. (ed.), *Karl Popper: Revisión de su legado*, Unión Editorial, Madrid, 2004.

GONZALEZ, W. J., "La evolución del Pensamiento de Popper," in GONZALEZ, W. J. (ed.), *Karl Popper: Revisión de su legado*, Unión Editorial, Madrid, 2004, pp. 23-194.

GONZALEZ, W. J. (ed.), *Science, Technology and Society: A Philosophical Perspective*, Netbiblo, A Coruña, 2005.

GONZALEZ, W. J., "The Relevance of Science, Technology and Society: The 'Social Turn'," in GONZALEZ, W. J. (ed.), *Science, Technology and Society: A Philosophical Perspective*, Netbiblo, A Coruña, 2005, pp. ix-x.

GONZALEZ, W. J., "The Philosophical Approach to Science, Technology and Society," in GONZALEZ, W. J. (ed.), *Science, Technology and Society: A Philosophical Perspective*, Netbiblo, A Coruña, 2005, pp. 3-49.

HACKING, I., *The Social Construction of* What*?*, Harvard University Press, Cambridge, MA, 1999.

HAHN, H., CARNAP, R. and NEURATH, O., *Wissenschaftliche Weltauffassung: Der Wiener Kreis*, Wolff, Vienna, 1929. Translated as: *The Scientific Conception of the World: The Vienna Circle*, Reidel, Dordrecht, 1973.

HARRE, R. (ed.), *Problems of Scientific Revolution: Progress and Obstacles to Progress in the Sciences*, Oxford University Press, Oxford, 1975.

HEMPEL, C. G., *Philosophy of Natural Science*, Prentice Hall, Englewood Cliffs, NJ, 1966.

HOWSON, C. and URBACH, P., *Scientific Reasoning: The Bayesian Approach*, Open Court, La Salle, IL, 1989 (reprinted in 1990). There is a second edition (Open Court, La Salle, IL, 1993), and a third one (2006).

IHDE, D., "Has the Philosophy of Technology Arrived? A State-of-the-Art Review," *Philosophy of Science*, v. 71, n. 1, (2004), pp. 117-131.

KITCHER, PH., "Arithmetic for the Millian," *Philosophical Studies*, v. 37, (1980), pp. 215-236.

KITCHER, PH., *The Nature of Mathematical Knowledge*, Oxford University Press, Oxford, 1983.

KITCHER, PH., "The Naturalist Returns," *Philosophical Review*, v. 101, (1992), pp. 53-114.

KITCHER, PH., *The Advancement of Science: Science without Legend, Objectivity without Illusions*, Oxford University Press, New York, 1993.

KITCHER, PH., *Science, Truth, and Democracy*, Oxford University Press, Oxford, 2001.

KNORR-CETINA, K., *The Manufacture of Knowledge. An Essay on the Constructivist and Contextual Nature of Science*, Pergamon Press, Oxford, 1981.

KOERTGE, N. (ed.), *A House Built on Sand: Exposing Postmodern Myths about Science*, Oxford University Press, N. York, 1998.

KUHN, TH. S., *The Structure of Scientific Revolutions*, The University of Chicago Press, Chicago, 1962 (2nd ed., 1970).

KUHN, TH. S., *The Road Since* Structure. *Philosophical Essays, 1970-1993, with an Autobiographical Interview*, edited by James Conant and John Haugeland, The University of Chicago Press, Chicago, 2000.

LAKATOS, I., *The Methodology of Scientific Research Programmes*, edited by John Worrall and Gregory Currie, Cambridge University Press, Cambridge, 1978.

LATOUR, B. and WOOLGAR, S., *Laboratory Life: The Social Construction of Scientific Facts*, Princeton University Press, Princeton, NJ, 1979; 2nd ed., Princeton University Press, Princeton, NJ, 1986.

LATOUR, B., *The Pasteurisation of France*, Harvard University Press, Cambridge, MA, 1988.

LATOUR, B., *We have Never been Modern,* Harvester, Brighton, 1993 (translated by C. Porter.)

LAUDAN, L., *Progress and its Problems. Towards a Theory of Scientific Growth*, University of California Press, Berkeley, 1977.

LAUDAN, L., "Some Problems Facing Intuitionistic Meta-Methodologies," *Synthese*, v. 67, (1986), pp. 115-129.

LAUDAN, L., "The Demise of the Demarcation Problem," in COHEN, R. and LAUDAN, L. (eds.), *Physics, Philosophy and Psychoanalysis*, Reidel, Dordrecht, 1983, pp. 111-128.

LAUDAN, L., "Progress or Rationality? The Prospect for Normative Naturalism," *American Philosophical Quarterly*, v. 24, n. 1, (1987), pp. 19-33.

LAUDAN, L., *Science and Relativism: Some Key Controversies in the Philosophy of Science*, The University of Chicago Press, Chicago, 1990.

LAUDAN, L., *Beyond Positivism and Relativism: Theory, Method, and Evidence*, Westview Press, Boulder, CO, 1996.

LAUDAN, L., "Naturalismo normativo y el progreso de la Filosofía," in GONZALEZ, W. J. (ed.), *El Pensamiento de L. Laudan. Relaciones entre Historia de la Ciencia y Filosofía de la Ciencia*, Publicaciones Universidad de A Coruña, A Coruña, 1998, pp. 105-116.

LAUDAN, L., "Una Teoría de la evaluación comparativa de teorías científicas," in GONZALEZ, W. J. (ed.), *El Pensamiento de L. Laudan. Relaciones entre Historia de la Ciencia y Filosofía de la Ciencia*, Publicaciones Universidad de A Coruña, A Coruña, 1998, pp. 155-169.

LAUDAN, L., "The Epistemic, the Cognitive, and the Social," in MACHAMER, P. and WOLTERS, G. (eds.), *Science, Values, and Objectivity*, University of Pittsburgh Press, Pittsburgh, and Universitätsverlag, Konstanz, 2004, pp. 14-23.

MACHAMER, P. and SILBERSTEIN, M. (eds.), *The Blackwell Guide to the Philosophy of Science*, Blackwell, Oxford, 2002.

MARTINEZ SOLANO, J. F., *El problema de la verdad en K. R. Popper: Reconstrucción histórico-sistemática*, Netbiblo, A Coruña, 2005.

NIINILUOTO, I. and TUOMELA, R., *Theoretical Concepts and Hypothetico-Inductive Inference*, Reidel, Dordrecht, 1973.

NIINILUOTO, I., "Scientific Progress," *Synthese*, v. 45, (1980), pp. 427-464 (reprinted in NIINILUOTO, I., *Is Science Progressive*?, Reidel, Dordrecht, 1984 pp. 75-110).

NIINILUOTO, I., *Is Science Progressive?*, Reidel, Dordrecht, 1984.

NIINILUOTO, I., "Progress, Realism and Verisimilitude," in WEINGARTNER, P. and SCHURZ, G. (eds.), *Logic, Philosophy of Science and Epistemology*, Holder-Pichler-Tempsky, Vienna, 1987, pp. 151-161.

NIINILUOTO, I., *Critical Scientific Realism*, Clarendon Press, Oxford, 1999, p. 10.

PITT, J. (ed.), *Change and Progress in Modern Science*, Reidel, Dordrecht, 1985.

POPPER, K. R., *Objective Knowledge. An Evolutionary Approach*, Clarendon Press, Oxford, 1972.

PUTNAM, H., "What Theories Are Not," in NAGEL, E., SUPPES, P. and TARSKI, A. (eds.), *Logic, Methodology and Philosophy of Science*, Stanford University Press, Stanford, 1962, pp. 240-251.

PUTNAM, H., "Meaning and Reference," *Journal of Philosophy*, v. 70, (1973), pp. 699-711. Reprinted in SCHWARTZ, S. P. (ed.), *Naming, Necessity, and Natural Kinds*, Cornell University Press, Ithaca, 1977, pp. 118-132.

PUTNAM, H., "Reference and Understanding," in PUTNAM, H., *Meaning and the Moral Sciences*, Routledge and K. Paul, London, 1978, pp. 97-119.

PUTNAM, H., "Reference and Truth," in PUTNAM, H., *Realism and Reason*, Cambridge University Press, Cambridge, 1983, pp. 69-86.

PUTNAM, H., *The Many Faces of Realism*, Open Court, La Salle, IL, 1987.

PUTNAM, H., *Realism with a Human Face*, Harvard University Press, Cambridge, MA, 1990.

QUINE, W. v. O., "Epistemology Naturalized," in QUINE, W. v. O., *Ontological Relativity and Other Essays*, Columbia University Press, N. York, 1969, pp. 69-90.

RADNITZKY, G. and ANDERSSON, G. (eds.), *Progress and Rationality in Science*, Reidel, Dordrecht, 1978.

RESCHER, N., *Scientific Progress: A Philosophical Essay on the Economics of Research in Natural Science*, University of Pittsburgh Press, Pittsburgh, 1978.

RESCHER, N., *Cognitive Economy: The Economic Perspectives of the Theory of Knowledge*, University of Pittsburgh Press, Pittsburgh, 1989.

RESCHER, N., *Priceless Knowledge? Natural Science in Economic Perspective*, University Press of America, Savage, MD, 1996.

RESCHER, N., *Objectivity: The Obligations of Impersonal Reason*, Notre Dame University Press, Notre Dame, IN, 1997.

RESCHER, N., "The Economic Dimension of Philosophical Inquiry," in RESCHER, N., *Philosophical Reasoning. A Study in the Methodology of Philosophizing*, B. Blackwell, Oxford, 2001, Ch. 8, pp. 103-115.

ROSENBERG, A., "A Field Guide to Recent Species of Naturalism," *The British Journal for the Philosophy of Science*, v. 47, (1996), pp. 1-29.

SALMON, W. C., "Rationality and Objectivity in Science, *or* Tom Kuhn Meets Tom Bayes," in SAVAGE, C. WADE (ed.), *Scientific Theories,* University of Minnesota Press, Minneapolis, 1990, pp. 175-204.

SCERRI, E. R. and WORRALL, J., "Prediction and the Periodic Table," *Studies in History and Philosophy of Science*, v. 32, n. 2, (2001), pp. 407-452.

SHÄFER, W. (ed.), *Finalization in Science: The Social Orientation of Scientific Progress*, Reidel, Dordrecht, 1983.

SCHARFF, R. C. and DUSEK, V. (eds.), *Philosophy and Technology: The Technological Condition*, Blackwell, Oxford, 2003.

SHRADER-FRECHETTE, K., "Objectivity and Professional Duties Regarding Science and Technology," in GONZALEZ, W. J. (ed.), *Science, Technology and Society: A Philosophical Perspective*, Netbiblo, A Coruña, 2005, pp. 51-79.

SIMON, H. A., *The Sciences of the Artificial*, 3rd ed., The MIT Press, Cambridge, MA, 1996.

STRAWSON, P. F., *Skepticism and Naturalism: Some Varieties*, Columbia University Press, New York, 1985.

SUPPE, F. (ed.), *The Structure of Scientific Theories*, University of Illinois Press, Urbana, 1974 (2nd. ed. 1977).

THAGARD, P., *Computational Philosophy of Science*, The MIT Press, Cambridge, MA, 1988.

THAGARD, P., *Conceptual Revolutions*, Princeton University Press, Princeton, 1992.

THAGARD, P., "Mind, Society and the Growth of Knowledge," *Philosophy of Science*, v. 61, (1994), pp. 629-645.

THAGARD, P., "Explaining Scientific Change: Integrating the Cognitive and the Social," in HULL, D., FORBES, M. and BURIAN, R. M. (eds.), *Proceedings of the 1994 Biennial Meeting of the Philosophy of Science Association*, Philosophy of Science Association, East Lansing, vol. 2, 1995, pp. 298-303.

WOOLGAR, S., "Critique and Criticism: Two Readings of Ethnomethodology," *Social Studies of Science*, v. 11, n. 4, (1981), pp. 504-514.

WOOLGAR, S., *Science: The Very Idea,* Tavistock, London, 1988.

WOOLGAR, S. (ed.), *Knowledge and Reflexivity: New Frontiers in the Sociology of Knowledge*, Sage, London, 1988.

WORRALL, J., "The Ways in Which the Methodology of Scientific Research Programmes Improves on Popper's Methodology," in RADNITZKY, G. and ANDERSSON, G. (eds.), *Progress and Rationality in Science*, Reidel, Dodrecht, 1978, pp. 45-70.

WORRALL, J., "Scientific Realism and Scientific Change," *Philosophical Quarterly*, v. 32, (1982), pp. 201-231.

WORRALL, J., "Scientific Discovery and Theory-Confirmation," in PITT, J. C. (ed.), *Change and Progress in Modern Science*, Reidel, Dordrecht, 1985, pp. 301-331.

WORRALL, J., "The Value of a Fixed Methodology," *The British Journal for the Philosophy of Science*, v. 39, (1988), pp. 263-275.

WORRALL, J., "Fix It and Be Damned: A Reply to Laudan," *The British Journal for the Philosophy of Science*, v. 40, (1989), pp. 376-388.

WORRALL, J., "Structural Realism: The Best of Both Worlds?," *Dialectica*, v. 43, n. 1-2, (1989), pp. 99-124.

WORRALL, J., "Fresnel, Poisson and the White Spot: The Role of Successful Predictions in the Acceptance of Scientific Theories," in GOODING, D., PINCH, T. and SCHAFFER, S. (eds.), *The Uses of Experiment: Studies of Experimentation in Natural Science*, Cambridge University Press, Cambridge, 1989, pp. 135-157.

WORRALL, J., "Rationality, Sociology and the Symmetry Thesis," *International Studies in the Philosophy of Science*, v. 4, n. 3, (1990), pp. 305-319.

WORRALL, J., "Scientific Revolutions and Scientific Rationality: The Case of the 'Elderly Holdout'," in SAVAGE, C. WADE (ed.), *Scientific Theories*, University of Minnesota Press, Minneapolis, 1990, pp. 319-354.

WORRALL, J., "Realismo, racionalidad y revoluciones," *Agora*, v. 17, n. 2, (1998), pp. 7-24.

WORRALL, J., "Programas de investigación y heurística positiva: Avance respecto de Lakatos," in GONZALEZ, W. J. (ed), *La Filosofía de Imre Lakatos: Evaluación de sus propuestas*, UNED, Madrid, 2001, pp. 247-268.

I

Methodological Approaches on Central Problems of Science

Scientific Reasoning and the Bayesian Interpretation of Probability

Colin Howson

1. Scientific Reasoning

The Bayesian theory supplies a *model* of inductive reasoning. When we want to theorise about a domain we often do so by constructing suitable models of it. There seems no good reason to prohibit reasoning itself from being the subject of this activity. Indeed, there already such models, and familiar ones. Systems of deductive logic are models of valid deductive reasoning. This does not mean that they are models of deductive reasoning as people actually do it, validly or not, but of valid deductive reasoning *in itself*. What is that? It is what all deductively valid inferences have in common by virtue of being deductively valid, whether anybody ever made or could make them. It is also what all deductively consistent sets of sentences, sets thought of and sets not thought of, and sets that in principle never could be thought of, have in common by virtue of being deductively consistent. What makes the theories we have of these things models is that they are constructed in a mathematically rather idealised way, over possible infinite sets of propositions closed under the Boolean combination, in "languages" that people would not generally think of as languages in any ordinary way, and employing "sharp" (truth or probability) values, which, besides being unrealistic, in the case of truth-values lead to Sorites-like paradoxes.

Despite their highly idealised character, such models are used to explain things, but I should emphasise the target is not what goes on in any "ideal reasoner's" head when they are reasoning: the task is to give a reason why appropriate samples of reasoning can be mapped, even approximately, into samples of what the model pronounces to be *valid* reasoning. This of course means that the model has to be *applicable*, which correspondingly means that it has to have a suitable amount of representative capacity —it has to be able to represent the sorts of *things* that human beings reason about, which for a model of inductive reasoning are things like data, hypotheses, etc. This we know the Bayesian model can do: it prescribes how the credibility of conjectures is enhanced, diminished or unaffected by evidence, good both in the sense of providing an illuminating explanatory model of scientific reasoning, and in showing how it is possible to make sound inductive inferences. In fact, of course, we should strictly be talking of Bayesian *models* in the plural, since for different purposes there are different ways of specifying the model: in some cases, where the mathematics is more subtle (as in the convergence-of-opinion theorems), the precise mathematical specification is important, in others, one can safely take a rough-and-ready approach to the mathematics. There are also large-scale and small-scale models, in which attention is confined to a class of possibilities describable within some restricted vocabulary, for example that describing the possible outcomes of an experiment, which in the extreme case can be just the labels '1' and '0' (standing for "as predicted" and "other than as predicted"), or sequences, possibly infinite sequences of these, corresponding to sequences of repetitions of such an experiment. For convenience I shall continue to talk of *the* Bayesian model; this can be regarded as defined in context.

What gives the Bayesian model its principal significance is that given suitable initial assumptions, not only can inductive reasoning be represented as valid reasoning, but the model(s) tell us that to reason validly in those circumstances *is necessarily to reason inductively*. This is surprisingly easy to show without going into the subtle technicalities of the convergence-of-opinion theorems, which are the results usually enlisted to show that the Bayesian theory has what one might call "inductive capability." One very simple consequence of the probability axioms is that if, $0 < P(E), P(H) < 1$ and H entails E modulo some initial conditions assigned probability 1 by P then $P(H|E) > P(H)$, i.e. H has its initial probability raised by E. All the "inductive theorems" state that, given suitable initial conditions, in this case distributions of prior probability and information about likelihoods, the Bayesian model explains inductive reasoning as a condition of maintaining the internal consistency of an agent's beliefs. Consistency with Humean inductive scepticism is maintained by noting that the inferences are valid, but depend on premises describing suitable belief-distributions. The situation is, *mutatis mutandis*, the same as that in deductive logic: you only get out some transformation of what you put in. Bacon notoriously castigated the contemporary version of deductive logic, Aristotelian syllogistic, for what he perceived as a crucial failing: deductive inference does not enlarge the stock of factual knowledge. But Bacon's condemnation has with the passage of centuries become modulated into a recognition of what is now regarded as a if not the fundamental property of logically valid inferences which, as we see, probabilistic reasoning shares, and in virtue of which it is an authentic logic: *valid inference does not beget new factual content*. Perhaps surprisingly in view of what is often said about it, valid probabilistic reasoning is therefore, just like deductive reasoning, *essentially non-ampliative*. I shall give a formal demonstration of this later, when I present the model explictly as a logic of inference sharing a close kinship with the deductive one.

Closer examination of the Bayesian model reveals a rich variety of inferential principles and inductive strategies. Detailed treatments exist in what has grown over the last quarter century into a very extensive literature. In addition, a variety of results has been proved which both extend the scope of the model and advance our understanding of the phenomena the theory deals with. Needless to say, this is reassuring both from the point of view of endorsing intuitively compelling strategies and from that of giving some indirect support to the model itself.

A noteworthy feature of all these model-explanations is that the assumptions which have to be made to generate the results figure explicitly in the reasoning: it is *transparent*. This is by contrast with most of what passes for inductive reasoning elsewhere, where many assumptions are implicit, sometimes with good reason since they would conflict with denials that any additional assumptions are in fact being made. We shall see some examples of this later, when we consider the well-known No-Miracles argument for scientific realism and the classic defence of significance testing in statistics, both formally very similar and both avowedly non-Bayesian. Another noteworthy feature of valid probabilistic arguments is that the assumptions which figure in an inference are represented somewhat differently than in the deductive case. There they figure as undischarged lines in a proof. The canonical form of an inductive assumption, on the other hand, is the assignment of an initial probability to a corresponding proposition or propositions, which will characteristically be employed with other such assumptions in the computation of some target probability. As we shall see later, however, this difference in the formal style of inference between the two cases, deductive and probabilistic, is actually more apparent than real. Before I come to that, let us look at the Bayesian model of scientific reasoning in a little more detail.

2. A Beginner's Guide to Bayesian Confirmation Theory

Considered independently of any particular context of application, the Bayesian model is a set of general principles —the probability axioms— governing the behaviour of a conditional probability function P(H|E); Bayesian confirmation theory is no more than the examination of its properties for various "interesting" instances of H and E. The most famous tool for analysing its behaviour of is contained in the eponymous consequence of the probability axioms, *Bayes's Theorem*. Here we shall be more interested in a particular rewriting of that theorem which displays P(H|E) explicitly as a function of P(H) and the *likelihood ratio*[1] $L = \mathrm{P}(\mathrm{E}|-\mathrm{H})/\mathrm{P}(\mathrm{E}|\mathrm{H})$:

$$\mathrm{P}(\mathrm{H}\,|\,\mathrm{E}) = \frac{\mathrm{P}(H)}{\mathrm{P}(H) + L\mathrm{P}(-\mathrm{H})} \qquad \text{(B)}$$

P(–H) of course is equal to 1 – P(H).

(B) is a remarkable formula, in terms of what it already reveals about the logic of inductive inference and what it can be made to reveal with a little work. In particular, the dependence of P(H|E) on the prior factor P(H) is a feature whose importance cannot be exaggerated, since it implies that *evidential data cannot by itself determine the credibility of a hypothesis*. This was obviously understood by Einstein when he wrote the following about Kaufmann's and Planck's experimental results concerning the velocity-dependence of the energy of a moving electron (these results appeared to agree with Abraham's theory, according to which the dynamics of electrons are based on purely electromagnetic considerations, and to be in conflict with the predictions of Special Relativity):

> "Herr Kaufmann has determined the relation between [electric and magnetic deflection] of β-rays with admirable care… Using an independent method, Herr Planck obtained results which fully agree with [the computations of] Kaufmann… It is further to be noted that the theories of Abraham and Bucherer yield curves which fit the observed curve considerably better than the curve obtained from relativity theory. However, in my opinion, these theories should be ascribed a *rather small probability* because their basic postulates concerning the mass of the moving electron are not *made plausible* by theoretical systems which encompass wider complexes of phenomena."[2]

It is not Einstein's specific reasons for assigning these probabilities that I want to emphasise here, but the fact that he is implicitly saying that if the prior probability is

---

1 Some authors call the reciprocal of L the likelihood ratio; it is also known as the *Bayes Factor*.

2 Translated and quoted in Pais, A., *Subtle is the Lord*, Clarendon Press, Oxford, 1982, p.159; my italics.

sufficiently large it can offset the weight of the empirical evidence; and in this explicit deference to the prior probability his inference is a characteristically Bayesian one.

Nevertheless, twentieth century philosophy of science was devoted, more or less, to denying that it made any sense to talk about the probabilities of hypotheses, and a fortiori to denying that the laws of epistemic probability furnished the logic of inductive inference: this was the view expressed in manifesto-like propouncements by the statisticians R. A. Fisher and later Neyman and Pearson, and the philosopher Karl Popper. According to these people all that is necessary for inductive inference in the theory of probability is knowledge of purely deductive relationships and of so-called "forward" probabilities, or *likelihoods* of the form $P(E|H)$. The latter merely state what the chance of E is *according to the hypothesis H itself.* If H entails E then $P(E|H) = 1$, while if H is a statistical hypothesis and E the sample data then $P(E|H)$ is the probability of that data according to the statistical model described by H. The fundamental idea is that a test of H is *severe* if it delivers some outcome E which is expected if H is true, and very unlikely if not. To quote Mayo, a modern exponent of this type of view, passing such a test can be legitimately regarded as furnishing "good grounds for H."[3]

We shall now see, using some simple examples, that this conclusion is incorrect. I shall start by examining a well-known defence of scientific realism. A favourite argument of realists for accepting that the relevant parts of it are "approximately" true is that the relatively established part of modern physical theory is established because it has successfully, often spectacularly, passed very severe tests in the sense above. For example, quantum electrodynamics predicts the observed value of the magnetic moment of the electron to better than one part in a billion, which would be highly improbable, to that same order of improbability, if the agreement were due merely to chance (whatever that means: I shall return to this point shortly), and there are predictions of general relativity which surpass even this extraordinary standard of success. The odds in these cases against the agreement being due simply to chance would therefore appear to be truly stupendous, and further multiplied if the rest of all the observational and experimental evidence is added in. And so we have the famous "No-Miracles Argument":

> "The positive argument for realism is that it is the only philosophy that doesn't make the success of science a miracle."[4]
>
> "*it cannot be just due to an improbable accident* if a hypothesis is again and again successful when tested in different circumstances, and especially if it is successful in making previously unexpected predictions … If a theory … has been well-corroborated, then it is highly probable *that it is truth-like*."[5]
>
> "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 the theory says about the fundamental structure of the

---

3 Cf. MAYO, D. G., *Error and the Growth of Experimental Knowledge,* The University of Chicago Press, Chicago, 1996, p. 177.

4 PUTNAM, H., *Mathematics, Matter and Method. Collected Papers,* vol. 1, Cambridge University Press, Cambridge, 1975, p. 73.

5 POPPER, K. R., *Realism and the Aim of Science,* Rowman and Littlefield, Totowa, 1983, p. 346; italics in original.

> universe being correct or ‘essentially’ or ‘basically’ correct. But we shouldn’t accept miracles, not at any rate if there is a non-miraculous alternative ... So it is plausible to conclude that presently accepted theories are indeed ‘essentially’ correct.”[6]

Just so that we are clear about the structure of this celebrated argument (it has a long pedigree going back three centuries) let us exhibit its essential features. They are as follows:

(1) Either current theory is “approximately” true, or agreement with the data is due to chance.

(2) If it is due to chance the probability of the data is exceptionally small.

(3) A very small probability for any hypothesis can be effectively set to zero if there is an alternative non-chance explanation.

(4) Current theory is such an alternative.

(5) *Therefore* current theory is approximately true.

However celebrated, the no-miracles argument is definitely unsound as it stands, for several reasons.[7] Firstly, it is no means clear what “chance” means as any sort of statable, definite hypothesis, to which can be added the problem of computing this ‘chance’ in any way that is not completely arbitrary (the computation, to the extent that one is ever provided, typically proceeds by applying the Principle of Indifference, but different choices of the possibility-space, or different *parameterisations* of a given space, will notoriously give different answers). Secondly, premise (1) is just false. It is a logical fact that there is an uncountable infinity of possible alternative explanations of the same data that do not involve any appeal to chance at all: there are, for a start, as Goodman told us, all the “grue_t” alternatives which add a time parameter to current theory and predict divergence at arbitrary later values of t. This is grist to the Bayesian’s mill, of course, since in the classical no-miracles argument it is clear that *these alternatives are implicitly discounted as carrying zero prior weight*. In other words, the no-miracles argument rests *essentially* on a suitable assignment of prior weights, independent of the data. Yet, ironically, the proponents of the no-miracles argument all claim to be opposed to the Bayesian apparatus of prior probabilities. But as we see, it is just these that drive the no-miracles argument.

It is no defence to claim that it is a far from trivial matter to produce any alternative that satisfies the sorts of stringent conditions that we currently think constrain the class of acceptable physical explanations (various symmetry conditions, correspondence principle, and so forth). It is no defence because if these constraints are sufficient to cut down the size of the class of “acceptable” alternatives then their content must itself extend far beyond the available empirical evidence. They are, in other words, extensionally equivalent to very strong *theories*, or classes of theories, *whose explicit acceptance is tantamount to assigning a prior probability of zero to all the alternatives inconsistent with them*. The “problem of priors” will not go away just because we choose to look the other way, and the apparatus of formal probability is nothing but a logical machinery for ensuring that we

---

[6] Worrall, J., “Fresnel, Poisson and the White Spot: The Role of Successful Predictions in the Acceptance of Scientific Theories,” in Gooding, D., Pinch, T. and Schaffer, S. (eds.), *The Uses of Experiment: Studies of Experimentation in Natural Science*, Cambridge University Press, Cambridge, 1989, p. 140; note the scare quotes.

[7] What follows is a condensed version of Howson, C., *Hume’s Problem*, Clarendon Press, Oxford, 2000, Chapter 3.

incorporate those prior evaluations in a transparent and consistent way (I shall justify the terminology "logical" later in this paper).

Thirdly, given that there are always non-chance alternatives to the hypothesis of agreement by chance (assuming that that has any definite meaning), premise (3) reduces simply to the injunction to regard small enough chances as in effect ignorable. This principle has a long pedigree in science: it goes back to Cournot, and is sometimes called *Cournot's Principle*, and it is the principle implicitly employed in the classic argument, due to R. A. Fisher, for rejecting the null hypothesis in the light of so-called significant sample evidence. According to Fisher the force of the rejection is

> "logically that of a simple disjunction: *Either* an exceptionally rare chance has occurred, *or* [the null hypothesis] is not true"[8]

with the implicit premise that exceptionally rare chances can be assumed not to occur, at any rate on this occasion. To this can be added the practical rider that if this strategy is adopted systematically in hypothesis-evaluation then a wrong rejection will be made only very rarely.

Lindley presents a graphic counterexample to this argument.[9] It assumes a typical textbook scenario: sampling from a specified normal distribution with unknown mean θ, where the null hypothesis H_0 is that θ takes a specified value θ_0. A sequence of observations $X = (x_1, \dots x_n)$ is determined from which a test statistic $T(X)$ is constructed.[10] $T(X)$ is supposed to carry all the information in X about θ, and large values of T are supposed to indicate increasing "discrepancy" with H_0. If there is no "directional" vague alternative to the null, T is taken to be the absolute value of the difference between the sample mean and θ_0, measured in units of the standard error, i.e. $T(X) = \bar{X} - \theta_0 | \sqrt{n} / \sigma$, where $\bar{X}$ is the mean of X. The P-value of a value t of T is the "tail-area" probability $P(T \geq t | H_0) = 2(1 - \Phi(t))$, where Φ is the standard normal distribution function. The P-value of t is supposed to measure the extent to which the discrepancy between X and θ_0 allegedly *contradicts* H_0. If the P-value of the observed value of T is α, and α is sufficiently small, then H_0 is correspondingly rejected "at that level of significance." Fisher explicitly used such quasi-deductive terminology, and in this he was followed by Popper who regarded small P-values as falsifiers of the corresponding null hypothesis.[11] According to Fisher again,

> "Tests of significance do not generally lead to any probability statements about the real world [i.e. to any form of Bayesian inference], but to a rational and well-defined measure of reluctance to the acceptance of the hypotheses they test."[12]

Fisher's explicit repudiation of any principle of probabilistic inference was, as we shall now see, unwise. Suppose that the observed value t of T is λ_α, i.e. the value of t such that $P(T \geq t | H_0) = \alpha$, and suppose that α is as small as you like. Note that because T is standardised λ_α does not depend on n, and hence as n takes larger values (though in

---

8 FISHER, R. A., *Statistical Methods and Statistical Inference*, Oliver and Boyd, London, 1956, p. 39.

9 Cf. LINDLEY, D. V., "A Statistical Paradox," *Biometrika*, v. 44, (1957), pp. 187-192.

10 I am following the standard practice of using upper-case to signify the random variable and lower-case its values.

11 Cf. POPPER, K. R., *The Logic of Scientific Discovery,* Hutchinson, London, 1959, p. 2003.

12 FISHER, R. A., *Statistical Methods and Statistical Inference*, p. 44.

any given experiment n is of course fixed) the same observation $T = \lambda_\alpha$ corresponds to values of $\bar{X}$ increasingly close to θ_0. Suppose also that a positive prior probability c, which can be arbitrarily small, is placed on H_0, with the remainder 1 – c spread uniformly over an interval I containing $\bar{\theta}$ (the latter assumption is for computational simplicity only; the discussion to follow does not depend on it). $\bar{X}$ is a sufficient statistic for θ and hence the posterior probability of H_0 given any value $\bar{X}$ is the same as it is given X. By Bayes's Theorem this posterior probability is

$$\frac{c\exp\left[-(1/2)\lambda_\alpha^2\right]}{c\exp\left[-(1/2)\lambda_\alpha^2\right]+(1-c)\,m^{-1}\int_I \exp\left[-(n/2\sigma^2)(\bar{x}-\theta)^2\right]d\theta}$$

where m is the length of I. For sufficiently large n, $\bar{x}$ will be well within I, since $\bar{x} - \theta_0$ tends to 0 (see (b) in the preceding paragraph), and so the integral is approximately $\sigma\sqrt{(2\pi/n)}$. Since λ_α does not depend on n, the quantity above tends to 1 as n increases. *In other words, for any* α *and c between 0 and 1 there is an n and an outcome* $\bar{x}$ *such that* $\bar{x}$ *is significant at level* ε *and the posterior probability of* H_0 *on* $\bar{x}$ *exceeds 1 –* ε. Lindley presented his example as a paradox in that

> "The usual interpretation of [the significance level] is that there is good reason to believe $\theta \neq \theta_0$; and of [the posterior probability] that there is good reason to believe $\theta = \theta_0$."[13]

But Lindley did not think the result was really a paradox. The fact is that any positive prior probability assignment to the null hypothesis is enough to make it overwhelmingly supported by a "significant" result, where the level of significance is as small as we please.

A closer analysis of this example is illuminating.[14] We can see from the Bayes's Theorem expression above (B) for the posterior probability that the likelihood ratio *L* (of probability densities, since we are dealing with continuous distributions) is approximately $\sigma\sqrt{(2\pi n^{-1})}$ / $\exp[(-1/2)\lambda_\alpha]^2$, which tends to 0 as n tends to infinity. It clearly has the same effect on the posterior probability as in the discrete case: the posterior probability is a decreasing function of *L*. This means that for a sufficiently small value of *L* the posterior probability will be as large as we like (given that the prior is nonzero, as it is in Lindley's example).

In the light of all these observations we can no see that the no-miracles argument is merely a probabilistic argument in which certain of the premises necessary to derive the conclusion are left unarticulated. Premise (1) becomes (1)': only the hypotheses of the approximate truth of current theory (H) and of chance (T) are worth assigning any positive prior probability to. Premises (2) and (3) can be combined into (2)': the probability of the data E conditional on T is exceptionally small compared with the probability of the data conditional on H. We can simply dispense with what we see is a fallacious rule to regard small probabilities as effectively impossibilities, since from (2)' we infer that the likelihood ratio $L = P(E|T) / P(E|H)$ is very small, and given (1)' we can set $P(-H) = P(T)$.

---

[13] LINDLEY, D. V., "A Statistical Paradox," p. 187.

[14] There is an extended discussion in HOWSON, C., "Bayesianism in Statistics," in SWINBURNE, R. (ed.), *Bayes's Theorem*, The British Academy, Oxford University Press, 2002, pp. 39-71.

Putting these values into (B) above we infer that if L is small enough, as by assumption it is, $P(H|E)$ is correspondingly close to 1. Q.E.D.

Whether you agree with all these amended premises is not the point here (because of the difficulty in endowing the chance hypothesis with a clear meaning from which any reliable value can be estimated for the probability in question, I myself do not). What is relevant is how indispensable the machinery of formal probability is to a valid reconstruction of it, and in general to the correct evaluation of hypotheses in the light of data. The latter conclusion is not uncontested, of course, the principal objection being the status of the prior distributions. These are subjective, it is said. But we already know this from the earlier discussion that some prior assessment has to be made, and that some sort of prior weighting is inevitable if any inference from the data is to be made. This doesn't always mean that the prior distributions always have to be worked out in any detail: for large enough samples the form of the prior is not very significant, with a uniform prior often being as good as any other. But that does not mean that the prior probabilities do not play an indispensable role. Nor are scientists, even eminent ones, averse from making such judgments explicitly: on the contrary, we have already seen Einstein doing just that, and such assessments occur very frequently in his published work. Here is another example:

> "I knew that the principle of the constancy of the velocity of light was something quite independent of the relativity postulate and I weighed which was the more probable… I decided in favour of the former."[15]

The fact is that, as we have seen above, some judgement of prior weighting is invariably made, explicitly or implicitly, and the task is to construct a suitable logic of inference developed within which these judgments are intergrated with the information obtained from the data. I claim that the Bayesian theory is the logic we require, and in the next section I shall try to justify that claim.

3. The Bayesian Interpretation of Probability

The claim certainly does require some justification. The usual ones, it must be said, fall a long way short of being ideal. That most favoured by contemporary Bayesians is, I believe, very deficient indeed. Due originally to F. P. Ramsey, and developed more systematically and elegantly by L. J. Savage, it derives the rules of epistemic probability from alleged rationality constraints on preferences. There are a number of problems attaching to specific parts of Savage's theory, among them some well-known paradoxical consequences of the Sure-Thing principle, as well as a serious question whether it is possible in that theory to separate questions of preference from judgments of probability.[16]

But there is also a more global feature of that approach which calls into question its claim to provide the foundation for a theory of rational belief, and that is the fact that its prescriptions are beyond the capacity of even a superhumanly rational believer to obey.

---

[15] Letter to Paul Ehrenfest, quoted in Stachel, J., *Einstein's Miraculous Year; Five Papers that Changed the Face of Physics*, Princeton University Press, Princeton, 1988, p.112, my italics.

[16] Cf. Schervish, M. J., Seidenfeld, T., and Kadane, J. B., "State-Dependent Utilities," *Journal of the American Statistical Association*, v. 85, (1990), pp. 840-847.

For according to Savage's theory this rational believer must have real-valued degrees of belief distributed over an *infinite* algebra of events (i.e. a set closed under the Boolean operations), which obeys among others the following rules:

- If $A \Rightarrow B$ then $P(A) \leq P(B)$
- $P(T) = 1$ where T is a logical truth,
- $P(\perp) = 0$ where $\perp$ is a contradiction.

But a celebrated result of Church and Turing implies that these prescriptions are, for any algebra of propositions based on even weakly expressive languages, *beyond the capacity even of a computer with unlimited storage and no physical constraints on the number of physical operations it can perform*. In the popular parlance, Savage's theory assumes "logically omniscient" agents.[17] Since no rational individual can satisfy these constraints (except by accident), it follows that this is not an acceptable theory of "the behaviour of a 'rational' person."[18] The objection is not substantially deflected by saying that Savage's theory is an acknowledged idealisation (whence the scare-quotes around 'rational' in the quotation). *Any* model of rational behaviour that makes no allowance for the fact that people are highly bounded reasoners, using irreducibly vague estimates of probability, is an inadequate model.

By contrast with this fanciful postulation of globally-defined belief functions satisfying impossibly restrictive assumptions, a more achievable and useful goal would be to show that the rules of probability are merely constraints on consistent assignments of probability, in much the same way that the clauses of a classical Tarskian truth-definition provide constraints on consistent truth-value assignments. In each case, an assignment is consistent just in case there exists a single-valued extension to the full algebra/language[19]. Note the "there exists" here: there is no presumption that the extension does or even can characterise anyone's state of belief. Such a consideration is quite irrelevant, as it should be. Indeed, one does not have to talk about anyone's beliefs at all: probability assignments are consistent (or not) independently of whether they are in fact made by anyone, just as truth-value assignments are consistent (or not) whether anyone actually makes them. It follows that according to this view, the Bayesian theory is not, as it is often claimed, an epistemology, but a logic very similar in scope and many of its formal properties to classical deductive logic.

I shall try to substantiate the claim to logical status shortly. First we must consider —and answer— the question by what right one can say that the probability axioms are rules of consistent assignments of probability. These rules are not obviously analytic of the informal idea of probability, even if the numbers are placed in the unit interval [0,1], and it merely begs the question to assert them without further explanation. There is (fortunately) an answer to the question, given independently by the American physicist R. T. Cox[20] and

---

[17] Cf. Earman, J., *Bayes or Bust? A Critical Examination of Bayesian Conformation Theory*, The MIT Press, Cambridge, MA, 1992, p. 124.

[18] Cf. Savage, L. J., *The Foundations of Statistics,* J. Wiley, New York, 1954, p. 7.

[19] The analogy with deductive consistency, a surprisingly close one, is developed in detail in Howson, C., "Probability and Logic," *Journal of Applied Logic*, v. 1, (2001), pp. 151-167.

[20] Cf. Cox, R.T., *The Algebra of Probable Inference*, The Johns Hopkins Press, Baltimore, 1961.

the English statistician and mathematician I. J. Good.[21] Cox and Good themselves agree that the probability axioms are not obviously analytic, and instead propose some very general rules which numerical assignments should obey *independently of any particular scale or method of measurement*. At this point I shall follow Cox's more developed account.[22] Using Cox's own notation (as he points out it derives from Keynes's classic work)[23], we suppose that A|C represents some acceptable method of assigning real-number-probabilities to propositional items of the form A given C, where A and C are propositions and C is deductively consistent, and the numbers are in any closed interval [a,b]. The following three principles seem fundamental

(i) –A|C should depend in some way on A|C, and decrease as A|C increases.

(ii) A&B|C should depend in some way on both A|B&C and B|C, and be an increasing function f of both.

(iii) These functional dependencies should be continuous and smooth, to the point of being twice-differentiable with continuous second derivatives.

These assumptions are in effect a system of functional equations to be solved for the two functions referred to, and Cox was able to solve them simultaneously. What he discovered was that the general solution can be expressed in the form

$$-A|C = h^{-1}[1 - h(A|C)]$$
$$A\&B|C = h^{-1}[h(A|B\&C) \times h(B|C)]$$

where h is a continuous, strictly increasing function from [a, b] to [0, 1] (hence h preserves ordering). It quickly follows that any acceptable method of assigning numbers to any set of propositions in a set closed under &, v and – can be rescaled (by such a rescaling function h) into the unit interval as formally a conditional probability function. The nub of Cox's proof is showing that the Boolean rules governing &, v and – entail that f is associative in its domain, i.e. that $f(x, f(y, z)) = f(f(x, y), z)$, a functional equation whose general solution can be expressed in the form $k\varphi(f(x,y)) = \varphi(x) \,.\, \varphi(y)$ where φ is strictly increasing but otherwise arbitrary.[24]

The Cox-Good result offers a novel and satisfying way of justifying the probability axioms: they are just a suitably scaled version of any acceptable way of assigning numerical probabilities. They are not, in other words, in themselves *absolute* consistency constraints, merely the consistency constraints appropriate to an admissible scale of measurement. This is just the result one should want: there is clearly nothing absolute about the ordinary scale of probability: it could in principle be transformed in any order-preserving way, with

---

[21] Cf. Good, I. J., *Probability and the Weighing of Evidence*, Griffin, London, 1950.

[22] Good's is a sketch contained in an Appendix to his book *Probability and the Weighing of Evidence*.

[23] Cf. Keynes, J. M., *A Treatise on Probability*, Macmillan, London, 1921.

[24] This part of Cox's proof has been criticised (notably by Halpern, cf. Halpern, J. Y., "Cox's Theorem Revisited," *Journal of Artificial Intelligence Research,* v. 11, (1999), pp. 429-435) because Cox does not show that the associativity condition is necessarily obeyed throughout [a,b]. Paris, who presents a proof of Cox's result which does not assume differentiability at all, adopts an explicit assumption that will guarantee this (Paris, J., *The Uncertain Reasoner's Companion*, Cambridge University Press, Cambridge, 1994, p. 24, assumption Co5), pointing out that the assumption restricts the scope of the result correspondingly. It seems to me that the criticism is unfair, and that by implicitly considering all possible values that the propositions might take, as well as those they actually do, Cox does ensure associativity everywhere in [a,b].

formally quite different rules for determining values on Boolean combinations. The choice of which particular scale to adopt, as Good and Cox point out, is largely conventional, and the usual one is chosen because of its familiar and simple properties.

We now have a general theory of consistency, subsuming both deductive and probabilistic, in which a consistent assignment, of truth-values in the one case and uncertainty values in the other, is one which is solvable subject to the relevant constraints. Which particular *logic* of consistency this determines will depend only on the nature of the values and the constraints. *Deductive logic is the logic of consistent assignments of truth-values, subject to the usual classical truth-definition constraints, while the logic of uncertainty is that of consistent assignments of probability, subject to the appropriate constraints*. The latter are such that they must be transformable into the usual rules of probability.

If we define the class of equivalent scales to be those which transform into each other by order-preserving continuous transformations, then the Cox-Good result shows that the admissible scales are equivalent to a probability scale. Let us choose the latter as the canonical representative of the admissible scales, and with it the (finitely additive) probability axioms. Now we can extend the theory of consistent assignments of probability to a definition of probabilistic consequence in exactly the same way as this is done in deductive (propositional) logic. Recall that a *model* of a consistent truth-valuation V(S) of a set S of propositional formulas is an extension of V(S) to a Boolean valuation V of all the formulas in the propositional language (a set closed under the usual truth-functional connectives). We can also say that an assignment V(T) to a set T is a *consequence* of V(S) just in case every model of V(S) is also a model of V(T) (this reduces in an obvious way to the more familiar definition of consequence for propositional sentences).

Similarly, suppose that L is an algebra of propositions (i.e. closed under the finitary Boolean operations) and S is a subset of L. Just as in propositional logic we define a *model* of a consistent uncertainty assignment V(S) (i.e. of numbers in [0, 1]) to be an extension of V(S) to a probability function on L, and we say that an assignment V(T) is a *consequence* of V(S) just in case every model of V(S) is a model of V(T). Thus V(T) is a consequence of V(S) just in case every probability function which extends V(S) assigns values to members of L as specified by V(T).

So we have a formally identical theory of semantic consistency and consequence for both deductive and probabilistic logic. Just as in deductive logic there are also systems of logical axioms for generating semantically valid consequences, so too there is here: the axioms of probability are now seen to be a system of logical axioms for generating valid probabilistic consequences from probabilistic premises. It can also be formalised as a type of sequent calculus just deductive logic can. Paris presents a formal deductive system II for sequents of the form K|K' (to be read "K entails K'"),[25] where K and K' are now finite sets of linear constraints (i.e. sets of the form $\Sigma u_{i,j} Bel(A_j) = v_i$, i = 1, … , m, where $Bel(A_j)$ is the agent's probability assignment to A_j and the sum is over j = 1, … , k, for some k, m, and where the axioms and rules of inference are explicitly specified. The A_i are sentences from some set SL of propositional sentences closed under the finitary Boolean operations. Paris then proves the following *completeness theorem* for this calculus:

[25] Cf. Paris, J., *The Uncertain Reasoner's Companion*, pp. 83-85.

> For any pair K, K', the set of all probability functions extending K to SL [i.e. the set of all models of K] is contained in those extending K' [i.e. the set of all models of K'] if and only if the sequent K|K' is a theorem of II.

A very important feature of the definition of deductive consistency is that if a truth-value assignment is consistent then it is so independently of the particular ambient propositional language. A theorem of de Finetti implies an analogous result for probabilistic consistency.[26] The local character of deductive consistency, and by implication deductively valid inference, is therefore completely mirrored in the account of probabilistic consistency and probabilistic consequence. This refutes the often-made but erroneous objection (see for example Cartwright)[27] that the Bayesian theory presupposes a fixed language over which all probabilities must be distributed *ab initio*. As we see, that is simply not true.

4. Coda: the Old-Evidence Problem

The Bayesian theory is supposed to reflect patterns of valid reasoning from data in terms of the way the data change one's probabilities. One type of such reasoning is assessing the impact of evidence on a hypothesis of data obtained *before* the hypothesis was first proposed. The stock example is the anomalous precession of Mercury's perihelion, discovered halfway through the nineteenth century and widely regarded as supporting Einstein's General Theory of Relativity (GTR) which was discovered (by Einstein himself) to predict it in 1915. Indeed, this prediction arguably did more to establish that theory and displace the classical theory of gravitation (CGT) than either of its other two dramatic contemporary predictions, namely the bending of light close to the sun and the gravitational red-shift. But according to nearly all commentators, starting with Glymour,[28] this is something which *in principle* the Bayesian theory cannot account for, since E is known then P(E) = 1 and it is a simple inference from the probability calculus that P(H|E) = P(H); i.e., such evidence cannot be a ground for changing one's belief in H.

Despite all this, the "old evidence" objection' is not a serious problem for the Bayesian theory; indeed, it is not a problem at all. What it really demonstrates is a failure to apply the Bayesian formulas sensibly, and to that extent the "problem" is rather analogous to inferring that $3/2 = x/x = 1$ from the fact that $3x = 2x$ if $x = 0$. To see clearly why we need only note an elementary fact about evidence, which is that *data do not constitute evidence for or against a hypothesis in isolation from a body of ambient information*. To talk about E being evidence relevant to H obviously requires a background of fact and information against which E is evaluated as evidence. A large dictionary found in the street is not in itself evidence either for

[26] Cf. Finetti, B. de, *Probability, Induction and Statistics,* John Wiley, London, 1972, p. 78.

[27] Cf. Cartwright, N., "Reply to Anderson," *Studies in History and Philosophy of Modern Physics,* v. 32, (2001), pp. 495-499. "Anderson urges Bayesian epistemology, which requires hypotheses and evidence to be pre-articulated. It is a poor choice for physics," Cartwright, N., "Reply to Anderson," p. 497. Contrast that remark with the fact that many prominent Bayesians have been physicists, and virtually all were or are working scientists. Here is a brief roll-call: Laplace (mathematician and physicist), Poincaré (mathematician and physicist), Harold Jeffreys (geophysicist), Frank Ramsey (mathematician, economist), Bruno de Finetti (mathematician and statistician), L. J. Savage (mathematician and statistician), Dennis Lindley (statistician), R. T. Cox (physicist), E. T. Jaynes (physicist), Abner Shimony (philosopher and physicist), Philip Anderson (Nobel laureate in physics).

[28] Cf. Glymour, C., *Theory and Evidence,* Princeton University Press, Princeton, NJ, 1980.

or against the hypothesis that Smith killed Jones. Relative to the background information that Jones was killed with a heavy object, that the dictionary belonged to Smith, and that blood found on the dictionary matches Jones's, it is. In other words, "being evidence for" connotes *a relation*, between the data, the hypothesis in question, *and* a body K of background information. The evidential weight of E in relation to H is assessed by how much E changes the credibility of H, in a positive or negative direction, given K.

Clearly, a condition of applying these obvious criteria is that K does not contain E. Otherwise, as the old-evidence "problem" reminds us, E could not *in principle* change the credibility of H. As Weisberg remarks

> "treating E as both the evidence and the background simultaneously is like evaluating the evidential import of a coin's having landed heads while using the fact that the coin landed heads as a background assumption. Such practice renders *all* evidence uninformative and is surely poor methodology."[29]

Indeed so.

We see that the old-evidence "problem," far from really being a problem, is merely an implicit reminder that if E is in K then it should first be deleted, as far as that can be done, before assessing its evidential weight. It is often objected against this so-called counterfactual strategy that there is no uniform method for deleting an item of information from a database K, and often it seems that there is no way at all which does not represent a fairly arbitrary decision. For example, the logical content of the set {A, B} is identical to that of $\{A, A \rightarrow B\}$, where A and B are contingent propositions, but simply subtracting A from each will leave two different sets of consequences; B will be in the first and not the second, for example, if the sets are consistent. Much has been made of this problem, and some have been led to believe that the task is hopeless. Fortunately, this is far from the truth. Suzuki has shown that there are consistent probabilistic contraction functions which represent the deletion of E from K relative to plausible boundary conditions on such functions (these conditions are furnished by the well-known AGM (Alchourrón-Gärdenfors-Makinson) theory of belief-revision).[30] The exhibition of a *particular* probability function representing the deletion of E will in general reflect the way the agent her/himself views the problem, and it is completely in line with the personalistic Bayesian theory adopted in this book that the request for an objective account of how this should be done is simply misplaced. Nevertheless, *it can usually be expected that the constraints imposed by the background information will practically determine the result*, and this is certainly true for the example which prompted the discussion, the observation of the precession of Mercury's perihelion, as I shall now show.

The discussion follows my book *Hume's Problem*.[31] We start, appropriately, with Bayes's Theorem, in the form (B) above, letting H be GTR, E the observed data on Mercury's perihelion (including the error bounds), p = P (H), but where P is like the agent's probability function *except that it does not "know" E*. We shall now show that, despite this

[29] WEISBERG, J., "Firing Squads and Fine-Tuning: Sober on the Design Argument," *The British Journal for the Philosophy of Science*, v. 56, (2005), p. 819.

[30] Cf. SUZUKI, S., "The Old Evidence Problem and AGM Theory," *Annals of the Japan Association for Philosophy of Science,* v. 1, (2005), pp. 1-20. See Suzuki's paper for references.

[31] Cf. HOWSON, C., *Hume's Problem*, p. 194.

idea sounding too vague to be useful, or even, possibly, consistent, *the data of the problem are sufficient to determine all the terms in (B) to within fairly tight bounds*.

Firstly, we have by assumption that H, together with the residual background information which P is assumed to "know," entails E, so $P(E|H) = 1$ by the probability axioms independently of any particular characteristic of P. Thus (B) becomes

$$P(H \mid E) = \frac{p}{p + P(E \mid -H)(1-p)}$$

and now we have only p and $P(E|-H)$ to consider. If we were to expand out $P(E|-H)$ we would find that it is a constant less than 1 multiplied by a sum whose terms are products $P(E|H_i)P(H_i)$, where H_i are alternatives to H. Recall now the assumption, reasonably appropriate to the situation in 1915, that the only serious alternative to GTR was CGT, meaning that it is the only H_i apart from H itself such that $P(H_i)$ is not negligible. Now we bring in the additional assumption that P does not "know" E. Judged on the residual background information alone, *the mere fact that E is anomalous relative to CGT means that $P(E|-H)$ will be very small*, say ε.

We are now almost there, with just p itself to evaluate. Remember that this too is evaluated on the residual background information. Without any of the confirming evidence for H, including E, this should mean that p, though small by comparison with P(CGT), which is correspondingly large (E is now not an anomaly for CGT, since by assumption E does not exist), is not negligible. It follows that, because of the very small likelihood ratio combined with a non-negligible if small prior probability, $P(H|E)$ can be made much larger than $p = P(H)$, and we see that H is correspondingly highly confirmed by E, *even though E is known*.

The "old evidence" problem arises from a misunderstanding of how to apply the Bayesian formalism properly, and to that extent it is analogous to declaring mathematics inconsistent because absurd answers are obtained from dividing by 0. Nor is it practically an issue: in every case that of interest the problem at hand should be a 'well posed' one, like the perihelion one above; i.e. the constraints should be sufficient to determine, up to sufficiently small errors, the terms in a Bayes's Theorem calculation. If the problem is not well-posed in this way, it is almost certainly of little or no practical interest anyway.

5. Bibliography

Cartwright, N., "Reply to Anderson," *Studies in History and Philosophy of Modern Physics*, v. 32, (2001), pp. 495-499.

Cox, R. T., *The Algebra of Probable Inference*, The Johns Hopkins Press, Baltimore, 1961.

Earman, J., *Bayes or Bust? A Critical Examination of Bayesian Conformation Theory*, The MIT Press, Cambridge, MA, 1992.

Finetti, B. de, *Probability, Induction and Statistics*, John Wiley, London, 1972.

Fisher, R. A., *Statistical Methods and Statistical Inference*, Oliver and Boyd, London, 1956.

Glymour, C., *Theory and Evidence*, Princeton University Press, Princeton, NJ, 1980.

GOOD, I. J., *Probability and the Weighing of Evidence*, Griffin, London, 1950.

HALPERN, J. Y., "Cox's Theorem Revisited," *Journal of Artificial Intelligence Research*, v. 11, (1999), pp. 429-435.

HOWSON, C. and URBACH, P. M., *Scientific Reasoning: the Bayesian Approach*, Open Court, Chicago, IL, 2nd edition, 1993.

HOWSON, C., *Hume's Problem*, Clarendon Press, Oxford, 2000.

HOWSON, C., "Probability and Logic," *Journal of Applied Logic*, v. 1, (2001), pp. 151-167.

HOWSON, C., "Bayesianism in Statistics," in SWINBURNE, R. (ed.), *Bayes's Theorem*, The British Academy, Oxford University Press, 2002, pp. 39-71.

KEYNES, J. M., *A Treatise on Probability*, Macmillan, London, 1921.

LINDLEY, D. V., "A Statistical Paradox," *Biometrika*, v. 44, (1957), pp. 187-192.

MAYO, D. G., *Error and the Growth of Experimental Knowledge*, The University of Chicago Press, Chicago, 1996.

PAIS, A., *Subtle is the Lord*, Clarendon Press, Oxford, 1982.

PARIS, J., *The Uncertain Reasoner's Companion*, Cambridge University Press, Cambridge, 1994.

POPPER, K. R., *The Logic of Scientific Discovery*, Hutchinson, London, 1959.

POPPER, K. R., *Realism and the Aim of Science*, Rowman and Littlefield, Totowa, 1983.

PUTNAM, H., *Mathematics, Matter and Method. Collected Papers*, vol. 1, Cambridge University Press, Cambridge, 1975.

SAVAGE, L. J., *The Foundations of Statistics*, Wiley, New York, 1954.

SCHERVISH, M. J., SEIDENFELD, T., and KADANE, J. B., "State-Dependent Utilities," *Journal of the American Statistical Association*, v. 85, (1990), pp. 840-847.

STACHEL, J., *Einstein's Miraculous Year; Five Papers that Changed the Face of Physics*, Princeton University Press, Princeton, 1988.

SUZUKI, S., "The Old Evidence Problem and AGM Theory," *Annals of the Japan Association for Philosophy of Science*, v. 1, (2005), pp. 1-20.

WEISBERG, J., "Firing Squads and Fine-Tuning: Sober on the Design Argument," *The British Journal for the Philosophy of Science*, v. 56, (2005), pp. 809-823.

WORRALL, J., "Fresnel, Poisson and the White Spot: The Role of Successful Predictions in the Acceptance of Scientific Theories," in GOODING, D., PINCH, T. and SCHAFFER, S. (eds.), *The Uses of Experiment: Studies of Experimentation in Natural Science*, Cambridge University Press, Cambridge, 1989, pp. 135-157.

Kuhn on Discovery and the Case of Penicillin

Donald Gillies

1. Kuhn's Theory of Discovery in Science

Kuhn is of course most famous for his theory of scientific revolutions. However in this paper I want to consider his theory of discovery in science. This theory is connected to his theory of scientific revolutions, but is nonetheless somewhat separate and very interesting in its own right. Kuhn first published his theory in a paper entitled: "The Historical Structure of Scientific Discovery." This appeared shortly before *The Structure of Scientific Revolutions* in the same year (1962). It has been reprinted in the collection *The Essential Tension* from where I will take my quotations. Much of the material in the paper was used in *The Structure of Scientific Revolutions*. It reappears largely in Chapter VI and Chapter X, page 114 of that work. I will base my account of Kuhn's theory both on his paper and his book.[1]

The plan of my own paper is as follows. I will begin in this section by expounding Kuhn's theory of discovery in science. I will then go on to describe a famous discovery which is not considered by Kuhn —namely the discovery of penicillin. I will give a historical account of this discovery in sections 2-4. Finally in section 5 I will consider how well Kuhn's theory fits this example. In some respects the fit is very good, and the example of the discovery of penicillin may be said to support some of Kuhn's general ideas on the subject of discovery very well. On the other hand the fit is not perfect and some modification of Kuhn's theory is needed to take account of the case of penicillin.

Kuhn is concerned to criticize the view of scientific discovery as: "a unitary event, one which, like seeing something, happens to an individual at a specifiable time and place."[2] Such an account he thinks applies at best to a relatively unproblematic kind of scientific discovery in which theory predicts a new sort of entity such as radio waves, and this entity is subsequently detected experimentally. There is, however, a different kind of scientific discovery which Kuhn refers to as "troublesome," and to which the account definitely does not apply. In such cases an entity is discovered which was not predicted by theory, and whose existence takes scientists by surprise. Often there seems to be an accidental element in such discoveries. This is how Kuhn himself makes the distinction:

> "The troublesome class consists of those discoveries —including oxygen, the electric current, X rays, and the electron— which could not be predicted from accepted theory in advance and which therefore caught the assembled profession by surprise… there is another sort and one which presents very few of the same problems. Into this second class of discoveries fall the neutrino, radio waves, and the elements which filled empty

---

[1] Cf. Kuhn, Th. S., "The Historical Structure of Scientific Discovery," *Science*, v. 136, (1962), pp. 760-764. Reprinted in Kuhn, Th. S., *The Essential Tension. Selected Studies in Scientific Tradition and Change*, The University of Chicago Press, Chicago, 1977, Chapter 7, pp. 165-177; and Kuhn, Th. S., *The Structure of Scientific Revolutions*, The University of Chicago Press, Chicago, 1962.

[2] Kuhn, Th. S., "The Historical Structure of Scientific Discovery," in Kuhn, Th. S., *The Essential Tension. Selected Studies in Scientific Tradition and Change*, p. 165.

> places in the periodic table. The existence of all these objects had been predicted from theory before they were discovered, and the men who made the discoveries therefore knew from the start what to look for."[3]

Kuhn's theory is concerned mainly with discoveries of the "troublesome" class. His main point is that such discoveries involve at least two steps, namely recognizing *that* something is, and recognizing *what* it is; or, to put the matter another way, observing something novel, and providing a theoretical explanation of that novelty. Because such discoveries are a complex process, they do not take place at an instant, and often more than one person is involved. As Kuhn says:

> "… discovering a new sort of phenomenon is necessarily a complex process which involves recognizing both *that* something is and *what* it is. Observation and conceptualization, fact and the assimilation of fact to theory, are inseparably linked in the discovery of scientific novelty. Inevitably, that process extends over time and may often involve a number of people. Only for discoveries in my second category —those whose nature is known in advance— can discovering *that* and discovering *what* occur together and in an instant."[4]

Kuhn illustrates his theory by the examples of the discovery of oxygen, of the planet Uranus, and of X rays. For this brief account of his views, I will confine myself to the example of the discovery of Uranus. Kuhn writes: "On the night of 13 March 1781, the astronomer William Herschel made the following entry in his journal: 'In the quartile near Zeta Tauri … is a curious either nebulous star or perhaps a comet.' That entry is generally said to record the discovery of the planet Uranus, but it cannot quite have done that."[5]

Indeed it cannot, because the discovery of Uranus was the discovery of a new planet unknown to previous astronomers. However, Herschel does not mention a planet, but speaks only of a "nebulous star or perhaps a comet." He does, however, observe that the object is "curious." Kuhn says:

> "On at least seventeen different occasions between 1690 and 1781, a number of astronomers, including several of Europe's most eminent observers, had seen a star in positions that we now suppose must have been occupied at the time by Uranus. One of the best observers in this group had actually seen the star on four successive nights in 1769 without noting the motion that could have suggested another identification. Herschel, when he first observed the same object twelve years later, did so with a much improved telescope of his own manufacture. As a result, he was able to notice an apparent disk-size that was at least unusual for stars."[6]

The key point here is that ordinary stars are not magnified in size by a telescope however powerful. They are so far away that they become effectively points of light rather than disks. The effect of a more powerful telescope is to make stars brighter and hence more visible rather than bigger. If a celestial object is increased in size by a telescope it must be something in the solar system such as a planet or comet, or a nebula. A nebula consists of

---

3 Kuhn, Th. S., "The Historical Structure of Scientific Discovery," pp. 166-167.

4 "The Historical Structure of Scientific Discovery," p. 171.

5 Kuhn, Th. S., "The Historical Structure of Scientific Discovery," p. 171.

6 Kuhn, Th. S., *The Structure of Scientific Revolutions*, p. 114.

a large collection of stars and so magnification in the sense of an increased separation of these stars becomes possible. Herschel would have been very familiar with this, and, as he observed a magnification of the star, he at once recognised this as "curious." He could have concluded that the object was a planet, a comet, or a nebula. In fact he rejected the correct one of these three possibilities and concluded that what he had seen was either a comet or a nebula. It was now easy to distinguish between these two possibilities. A comet would move against the background of the stars, while a nebula would remain fixed. Herschel observed his curious object on two further occasions, namely 17 and 19 March. The object moved, and he therefore concluded that it must be a comet. So he announced to the scientific community that he had discovered a new comet. Kuhn now continues the story as follows:

> "… astronomers throughout Europe were informed of the discovery, and the mathematicians among them began to compute the new comet's orbit. Only several months later, after all those attempts had repeatedly failed to square with observation, did the astronomer Lexell suggest that the object observed by Herschel might be a planet. And only when additional computations, using a planet's rather than a comet's orbit, proved reconcilable with observation was that suggestion generally accepted. At what point during 1781 do we want to say that the planet Uranus was discovered? And are we entirely and unequivocally clear that it was Herschel rather than Lexell who discovered it?"[7]

The example of the discovery of Uranus illustrates perfectly the two main claims which Kuhn makes about discoveries of his "troublesome" class. These are that (1) such a discovery is "necessarily a complex process which involves recognizing both *that* something is and *what* it is", and (2) that "inevitably, that process extends over time and may often involve a number of people." The main features of the story seem to be the following. First of all Uranus was seen by astronomers no less than seventeen times before Herschel's crucial observation, but none of these astronomers realised that there was anything unusual about what they had seen. Herschel, in contrast to these predecessors, did realise that he had seen something which, in his own words, was curious. This is an example of what Kuhn refers to as "the individual skill, wit, or genius to recognize that something has gone wrong in ways that may prove consequential."[8] But Kuhn also points out that Herschel's success was the result not just of greater "wit or genius," but depended on his having a better telescope. We can add that background knowledge also played a crucial part here. Herschel's better telescope enabled the magnification in size of Uranus to become more obvious, but it was his background knowledge which enabled him to realise that this was significant. An amateur without this background knowledge would not have made the discovery. But although Herschel realised that there was something unusual, he misinterpreted what he had seen as a comet. Thus the discovery was only complete when Lexell, as a result of elaborate theoretical calculations, did finally identify Uranus as a planet. So the discovery of Uranus did definitely extend over time, and did involve more than one person. We could perhaps say that Herschel discovered *that* there was something new and of interest, while Lexell discovered *what* it was. The example is

---

7 Kuhn, Th. S., "The Historical Structure of Scientific Discovery," p. 172.

8 "The Historical Structure of Scientific Discovery," p. 173.

also in agreement with the ideas of Fleck who was a major influence on Kuhn. Fleck in fact writes:

> "If any discovery is to be made accessible to investigation, the *social point of view* must be adopted; that is, the discovery must be regarded as a *social event*."[9]

This completes my account of Kuhn's theory of discoveries of the "troublesome" class. I will now begin my examination of whether the discovery of penicillin is in accordance with Kuhn's model. The first point to notice is that the discovery of penicillin agrees with Kuhn in that it took place over a quite long period of time, and involved several people. We may in fact distinguish two phases in the discovery. The first phase was the one which involved Alexander Fleming. Fleming, as the result of a chance observation of a contaminated Petri dish, discovered the existence of a new substance which powerfully inhibited a variety of pathogenic bacteria. Indeed some might regard this as constituting the discovery of penicillin. However, to me it does not seem that this in itself was all that there was to the discovery, because Fleming did not discover that penicillin could be used as what we would now call an antibiotic. In fact he had reasons for supposing that penicillin would not work as an antibiotic, and he himself used penicillin for another purpose. This brings me to the second phase in the discovery of penicillin which involved Howard Florey and his team at Oxford. They were the ones who showed that the substance which Fleming had discovered could be used successfully as an antibiotic. It will be seen that there is an analogy here to the contributions which Herschel and Lexell made to the discovery of Uranus. However, I will return to philosophical analysis in section 5 of the paper. In the next 3 sections (2, 3, and 4), I will give a brief historical account of how penicillin was discovered.[10]

2. The Discovery of Penicillin Phase 1: Fleming's Work

It was early in September 1928 that Fleming noticed an experimental plate in his laboratory which had been contaminated with a penicillium mould. If, however, we are to understand his reaction to this fateful event, we must first examine some of the research which Fleming had carried out previously. There were in fact two episodes which had influenced Fleming in a crucial fashion. The first of these was Fleming's experiences in the First World War, and this will be described in section 2.1. The second was Fleming's discovery in 1921 of an important biochemical substance which was named *lysozyme*. This will be dealt with in section 2.2. An understanding of how these two episodes had prepared Fleming's mind will enable us to understand why Fleming acted as he did when he stumbled on penicillin itself, and this will be described in section 2.3.

---

[9] Fleck, L., *Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Einführung in die Lehre von Denkstill und Denkkollektiv*, Schwabe, Basil, 1935. Translation by Fred Bradley and Thaddeus J. Trenn: *Genesis and Development of a Scientific Fact*, The University of Chicago Press, Chicago, 1979, p. 76.

[10] My account of the discovery of penicillin is largely based on Hare, R., *The Birth of Penicillin and the Disarming of Microbes*, George Allen and Unwin, London, 1970; MacFarlane, G., *Howard Florey. The Making of a Great Scientist*, Oxford University Press, Oxford, 1979, and MacFarlane, G., *Alexander Fleming. The Man and the Myth*, Chatto and Windus, The Hogarth Press, 1984. Hare's book is partly an eye witness account since he was working in the same laboratory as Fleming when Fleming made his discovery. Macfarlane's two books are excellent historical works which are informed by a deep scientific knowledge of the area.

2.1. Fleming's Experiences in the First World War

Fleming spent most of his career carrying out research in bacteriology in the inoculation department of St Mary's Hospital, London. This department was headed until his retirement in 1946 by Sir Almroth Wright. When the First World War broke out in 1914, Wright, Fleming and the rest of the department were sent to Boulogne to deal with the war wounded, and, in particular, to try to discover the best way of treating infected wounds. At that time wounds were routinely filled with powerful antiseptics which were known to kill bacteria outside the body. Fleming, however, soon made the remarkable discovery that bacteria seemed to flourish in wounds treated with antiseptics even more than they did in untreated wounds. The explanation of this apparent paradox was quite simple. In an untreated wound the bacteria causing the infection were attacked by the body's natural defences, the white cells, or *phagocytes*, which ingested the invading bacteria. If the wound was treated with an antiseptic, some bacteria were indeed killed, but the protective phagocytes were also killed, so that the net effect was to make the situation worse than before. Wright and his group therefore maintained (quite correctly) that wounds should not be treated with antiseptics. They advocated the earliest possible surgical removal of all dead tissue, dirt, foreign bodies, and so forth, and the irrigating the wound with strong sterile salt solution. The medical establishment of the day rejected this recommendation, and so the superior treatment was accorded only to those directly in the care of Wright and his team.

2.2. The Discovery of Lysozyme

After the war, Fleming returned to the inoculation department in London, and here in 1921 he discovered an interesting substance which was given the name *lysozyme*. Lysozyme was capable of destroying a considerable range of bacteria, and was found to occur in a variety of tissues and natural secretions. Fleming first came across lysozyme while studying a plate-culture of some mucus which he took from his nose when he had a cold. He later discovered that lysozyme is to be found in tears, saliva, and sputum, as well as in mucus secretions. He extended his search quite widely in the animal and vegetable kingdoms, and found lysozyme in fish eggs, birds' eggs, flowers, plants, vegetables, and the tears of more than fifty species of animals. Lysozyme destroyed about 75% of the 104 strains of airborne bacteria and some other bacteria as well. Moreover, Fleming was able to show that, unlike chemical antiseptics, even the strongest preparations of lysozyme had no adverse effects on living phagocytes, which continued their work of ingesting bacteria just as before. From all this, it seemed that lysozyme was part of many organisms' natural defence mechanisms against bacterial infection. Lysozyme had only one drawback. It did not destroy any of the bacteria responsible for the most serious infections and diseases. The hypothesis naturally suggested itself that the pathogenic bacteria were pathogenic partly because of their resistance to lysozyme.

If we put together Fleming's research on war wounds and his research on lysozyme, a problem situation emerges which I will call the "antiseptic problem situation." On the one hand, the chemical antiseptics killed pathogenic bacteria outside the body, but were less effective for infected wounds, partly because they destroyed the phagocytes as well. On the other hand, the naturally occurring antiseptic lysozyme did not kill the phagocytes, but also failed to destroy the most important pathogenic bacteria. The problem then was to discover a "perfect antiseptic" which would kill some important pathogenic bacteria without affecting the phagocytes. The work on lysozyme suggested that such antiseptics might be produced by naturally occurring organisms.

2.3. Fleming Stumbles on Penicillin

This then is the background to Fleming's work on penicillin. Fleming actually stumbled on penicillin while he was carrying out a fairly routine investigation. He had been invited to contribute a section on the staphylococcus group of bacteria for the nine-volume *A System of Bacteriology* which was being produced by the Medical Research Council. Fleming's contribution did indeed appear in the second volume in 1929. Staphylococci are spherical bacteria which are responsible for a variety of infections. For example, the golden-coloured *Staphylococcus aureus* is responsible for skin infections such as boils and carbuncles, as well as for a variety of other diseases. While reading the literature on staphylococci, Fleming came across an article by Bigger, Boland, and O'Meara of Trinity College, Dublin, in which it was suggested that colour changes took place if cultures of staphylococci were kept at room temperature for several days. This interested Fleming, because the colour of a staphylococcus can be an indicator of its virulence in causing disease. He therefore decided to carry out an experimental investigation of the matter with the help of D. M. Pryce, a research scholar.

The staphylococci were cultured in glass dishes, usually 85 mm in diameter, known as *Petri dishes*. These dishes were filled with a thin layer of gelatinous substance called *agar* to which enough nutrients could be added to allow the microbes to multiply. Using a platinum wire, some staphylococci were spread across the surface of the agar, and the plate was then incubated at a suitable temperature (usually 37ºC), to allow the microbes to multiply. After this period of incubation, the dish was set aside on the bench, and was examined every few days to see if changes in the colour of some of the staphylococci could be observed.

While this investigation was continuing, Pryce left the laboratory in February 1928 to start another job, but Fleming continued the work on his own throughout the summer. At the end of July Fleming went off for his usual summer holiday, leaving a number of culture-plates piled at the end of the bench where they would be out of the sunlight. Early in September (probably on 3 September) when Fleming had returned from his holiday, Pryce dropped in to see him. Pryce found Fleming sorting out the pile of plates on his bench. Discarded plates were put in a shallow tray containing the antiseptic Lysol. This would kill the bacteria, and make the Petri dishes safe for the technicians to wash and prepare for use again. Fleming's tray was piled so high with dishes that some of them were protruding above the level of the Lysol. Fleming started complaining about the amount of work he had had to do since Pryce had left him. He then selected a few of the dishes to show to Pryce. More or less by chance he picked up one in the tray of discards but above the level of the Lysol. According to Pryce's later recollection, Fleming looked at the plate for a while, and then said: "That's funny." The plate was in fact the famous penicillin plate.

This is how Fleming himself described what happened in the paper which he published in June 1929: "While working with staphylococcus variants a number of culture-plates were set aside on the laboratory bench and examined from time to time. In the examinations these plates were necessarily exposed to the air and they became contaminated with various micro-organisms. It was noticed that around a large colony of a contaminating mould the staphylococcus colonies became transparent and were obviously undergoing lysis."[11]

Fleming's photograph of the original penicillin plate is reproduced in Plate 1, and it is easy to follow his description when examining the photograph.

The colonies of staphylococci are the small circular blobs, and the contaminating mould is very obvious. Near the mould the staphylococci become transparent or disappear altogether. They are obviously, as Fleming says, undergoing *lysis*, which means the dissolution of cells or bacteria. From his observation of the plate, Fleming inferred that the mould was producing a substance capable of dissolving bacteria. The mould was identified as being a *Penicillium*. At first it was incorrectly classified as *Penicillium rubrum*, but later it was found to be the much rarer species *Penicillium notatum*. Fleming accordingly gave the name *penicillin* to the bacteriolytic substance which he thought was being produced by the mould.

The events described so far may make it look as if Fleming's discovery was simply a matter of luck. Indeed, there is no doubt that a lot of luck was involved. Hare subsequently tried to reproduce a plate similar to Fleming's original one, and found to his surprise that it was quite difficult.[12] The general effect of Fleming's plate could be produced only if the mould and the staphylococci were allowed to develop at rather a low temperature. Even room temperature in the summer would usually be too high, but here the vagaries of the English weather played their part. By examining the weather records at Kew, Hare discovered that for nine days after 28 July 1928 (just when Fleming had gone on holiday!), there was a spell of exceptionally cold weather. In addition to this, Hare concluded that

---

[11] Fleming, A., "On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of *B. Influenzae*," *British Journal of Experimental Pathology*, v. 10, (1929), p. 226.

[12] See Hare, R., *The Birth of Penicillin and the Disarming of Microbes*, pp. 66-80.

"… the plate cannot have been incubated in the usual way."[13] A final point is that the strain of penicillium which contaminated Fleming's plate is a very rare variety, and most penicillia do not produce penicillin in sufficient quantity to give rise to the effect which Fleming observed. How did such a rare mould find its way into Fleming's laboratory? The most likely explanation is a curious one. There was at that time a theory that asthma was caused by moulds growing in the basements of the houses in which the asthmatics lived. This theory was being investigated by the scientist C. J. La Touche in the laboratory immediately below Fleming's, and La Touche had as a result a large collection of moulds taken from the houses of asthma sufferers. It seems probable that *penicillium notatum* was one of these moulds.

There is no doubt then that a great deal of luck was involved in the discovery of penicillin. Yet it still needed creativity and insight on Fleming's part to seize the opportunity which chance had presented to him. Nothing shows this more clearly than a comparison of Fleming's reaction to the contaminated plate with that of his colleagues in the laboratory (including the head of the laboratory, Sir Almroth Wright) when he showed it to them. With characteristic candour, Hare describes the complete lack of interest shown by himself and the others:

> "The rest of us, being engaged in researches that seemed far more important than a contaminated culture plate, merely glanced at it, thought that it was no more than another wonder of nature that Fleming seemed to be forever unearthing, and promptly forgot all about it.
>
> The plate was also shown to Wright when he arrived in the afternoon. What he said, I do not recollect, but … one can assume that he was no more enthusiastic —he could not have been less— than the rest of us had been that morning."[14]

Fleming was by no means discouraged by his colleagues' cool reaction. He took a minute sample of the contaminating mould, and started cultivating it in a tube of liquid medium. At some later stage he photographed the plate, and made it permanent by exposing it to formalin vapour, which killed and fixed both the bacteria and the mould. Fleming kept the plate carefully, and it is now preserved in the British Museum. The whole episode then is a perfect instance of the famous claim made by Pasteur in his inaugural lecture as professor at Lille in 1854 when he said that: "In the field of observation fortune favours only the prepared mind."[15] Let us now examine how Fleming's mind had been prepared to appreciate the significance of his contaminated culture plate.

It is interesting in this context to compare Fleming with Herschel. Herschel needed a prepared mind to realise that a celestial body which was magnified by his telescope was something unusual. Specifically he needed the background knowledge that ordinary stars were not magnified by a telescope but only made brighter, and that, if something was magnified by a telescope, it had to be either an object within the solar system or a nebula. However, these bits of knowledge would have been part of the background of

---

13 Hare, R., *The Birth of Penicillin and the Disarming of Microbes*, p. 79.

14 Hare, R., *The Birth of Penicillin and the Disarming of Microbes*, p. 55.

15 In Pasteur's original French, the quotation runs: "Dans les champs de l'observation le hasard ne favorise que les esprits préparés." This is slightly ambiguous since "le hasard' in French can mean 'luck or fortune' as it is translated here, or 'chance' in the statistical sense.

any competent astronomer. The background knowledge which made Fleming appreciate the significance of the penicillin plate was, by contrast, rather unusual since it was not possessed by his very competent colleagues. It is not difficult, however, to see how this background knowledge which was specific to Fleming arose from his earlier researches.

Fleming, during his researches on lysozyme, had over and over again observed a substance produced from some naturally occurring organism destroying bacteria. This was a phenomenon with which he was very familiar. However, it would have struck him immediately that there was something new and curious about the penicillin case because the bacteria being inhibited were pathogenic staphylococci, whereas lysozyme only destroyed non-pathogenic bacteria. From the time of his work on the healing of wounds in the first world war, Fleming had been aware of the problem of finding a "perfect antiseptic" —that is an antiseptic which would kill pathogenic bacteria without destroying the phagocytes. In the light of his knowledge of this problem, it is reasonable to suppose that, when he saw the penicillin plate, he conjectured that the mould might be producing a "perfect antiseptic."

The assumption that Fleming made such a conjecture is borne out by his subsequent actions. Fleming grew the mould on the surface of a meat broth, and then filtered off the mould to produce what he called "mould juice." He then tested the effect of this mould juice on a number of pathogenic bacteria. The results were encouraging. The virulent streptococcus, staphylococcus, pneumococcus, gonococcus, meningococcus, and diphtheria bacillus were all powerfully inhibited. In fact, mould juice was a more powerful germicide than carbolic acid. At the same time, mould juice had no ill effects on phagocytes. Here at last seemed to be a "perfect antiseptic." Indeed in his 1929 paper, Fleming wrote: "It is suggested that it [penicillin] may be an efficient antiseptic for application to, or injection into, areas infected with penicillin-sensitive microbes."[16]

However, at this point a series of difficulties began to emerge for Fleming and his colleagues who were working on penicillin. These difficulties led Fleming to the conclusion that penicillin would not after all be the kind of "perfect antiseptic" which he had been hoping to find. He did, however, find another important use for penicillin. There is a certain analogy here to Herschel who identified his curious heavenly body as a comet rather than a planet. This is why Herschel's work was only the first phase in the discovery of Uranus, and a second phase carried out by Lexell was needed to establish the existence of a hitherto unknown planet. In the same way Fleming's work was only the first phase in the discovery of penicillin, and a second phase carried out by Florey and his Oxford team was needed to establish that penicillin was after all a "perfect antiseptic" —what we would now call a "powerful antibiotic." In the next section I will discuss the reasons which led Fleming and his team to give up the conjecture that penicillin would be a perfect antiseptic.

3. Why Fleming Abandoned his Hope that Penicillin would be a "Perfect Antiseptic"

Fleming did not leave behind a diary or detailed notebook setting out the reasons behind his changes in research strategy, so that these reasons have to be inferred from what he did, and naturally this can lead to differences in opinion. There are, however, three factors

[16] Fleming, A., "On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of *B. Influenzae*," p. 236.

which most historians would agree might have influenced Fleming in abandoning his early attempts to demonstrate that penicillin was a powerful antibiotic. The first of these was the fact that some of the results of tests carried out by Fleming and his collaborators seemed to indicate that penicillin would not work against bacteria when injected into the body. In section 3.1, I will discuss these "counter indications." A second factor was that there were considerable difficulties in isolating penicillin from mould juice. These problems will be considered in section 3.2. The third factor was that neither Fleming nor any of his colleagues carried out what is known as an animal protection tests. By contrast, tests of this kind were done by Florey and his Oxford team. This issue will be discussed in section 3.3. We now come to another surprising twist in the story of the development of penicillin. Although Fleming seems to have given up his initial hope that penicillin might be a perfect antiseptic, he did not abandon penicillin altogether because he found another use for it. In section 3.4 I will explain what this use was, and why it had a very positive influence on the further development of research into penicillin.

3.1. The Counter-Indications

Although the results of Fleming's first tests were encouraging, further experiments gave reasons to doubt whether penicillin would be an effective systemic antibacterial agent. First of all Fleming discovered that while chemical antiseptics killed microbes in a few minutes, his mould juice took several hours to do so. Then on 22 March 1929, Fleming's collaborator Craddock injected 20cc of penicillin into the ear vein of a rabbit. 30 minutes later a blood sample showed that almost all penicillin activity had disappeared. So if penicillin required about 4 hours to kill bacteria, but had disappeared 30 minutes after injection, it looked as if it could not work. Another finding of Craddock made the situation look even worse. Craddock discovered that penicillin lost about 75% of its activity in mixtures containing blood serum. Now as there is a great deal of serum in infected areas, this again strongly suggested that penicillin would not work as a "perfect antiseptic" if injected into the body.

So we see that Fleming had good reasons for giving up his "perfect antiseptic" hypothesis, but was he too Popperian in doing so? This example perhaps shows that Popper is too fiercely insistent on the need for scientists to give up hypotheses which appear to have been refuted.

3.2. Difficulties of Isolation

Another problem facing Fleming was that of isolating the active ingredient (penicillin) from mould juice. Fleming was a bacteriologist not a chemist, and it could be argued that the chemical problems of isolating and storing penicillin were what caused him to abandon his research on it. This theory seems to me false, however, because three skilful chemists worked in Fleming's laboratory around this time, namely Craddock, Ridley and Holt. As we shall see, between them they took most of the key steps for the extraction of penicillin which were later carried out by the Oxford team. This leads me to think that, if Fleming had retained his belief that penicillin might be a "perfect antiseptic," the chemical difficulties of extraction could have been overcome. However, he is unlikely to have thought it would be worth taking a lot of time and trouble to extract something which would not work. In other words, the counter-indications probably influenced Fleming more than the chemical difficulties of extracting penicillin from mould juice.

3.3. Absence of an Animal Protection Test

There is another important factor connected with Fleming's early work on penicillin. Neither he nor his collaborators ever performed an animal protection test. This is a test in which an animal, e.g. a mouse, is infected, and then injected with the drug being investigated to see if it cures the animal. Craddock, as we have seen, carried out an experiment on a rabbit, but this was not an animal protection test in the sense just defined.

It is in this connection that the discovery of the sulphonamide drugs in 1935 was very important for the further development of penicillin. These drugs were discovered in Germany as a by-product of the activities of the giant chemical company I. G. Farben. The discovery was made by a team headed by Gerhard Domagk, who was born in 1895 and appointed at the early age of thirty-two as director of research in experimental pathology and bacteriology in the institute attached to the I. G. Farben works at Elberfeld. Domagk and his team had huge laboratories in which they routinely tested compounds produced by the firm's industrial chemists on thousands of infected animals to see if the compounds had any therapeutic value.

The I. G. Farben chemists H. Hoerlein, W. Dressel, and F. Kothe produced a rich red dye which was very effective with protein materials such as wool and silk. This was known as *prontosil rubrum*. Domagk and his team then discovered that this same compound possessed the definite ability to cure mice infected with haemolytic streptococci. Domagk published this finding in 1935, but referred back to experiments carried out in 1932.

Now the interesting thing about this case is that the pharmaceutical value of prontosil rubrum could not have been discovered without the use of animal protection tests for the simple reason that prontosil rubrum does not inhibit bacteria in Petri dishes (*in vitro*). It is only when prontosil rubrum is injected into living creatures (used *in vivo*) that it acts as an effective agent against bacteria.[17] This suggested that penicillin too, despite the discouraging *in vitro* results, might work *in vivo*.

A personal difference between Fleming and Florey may also have been important here. Fleming was the deputy of Almroth Wright who cast scorn on random experiments of the I. G. Farben type and argued that a good scientist should proceed by making deductions from a few carefully chosen tests. There is much to be said for Almroth Wright's approach, but, in this instance, it led to the wrong conclusion. Moreover, Almroth Wright and Fleming almost never conducted animal experiments. They worked *in vitro*, but not *in vivo*. Florey, on the other hand, had been trained at working on physiology through animal experiments, and was very skilled at experimental animal surgery. For him, animal experiments were a matter of routine.

3.4. Why Fleming Nonetheless Preserved the Penicillin Mould

Despite abandoning his hopes that penicillin might be a "perfect antiseptic," Fleming nonetheless continued the cultivation of the penicillin mould and the production of mould juice. This was extremely fortunate since the mould (*penicillium notatum*) was a very rare

[17] Further details about the discovery of prontosil rubrum, and the explanation of why it works only *in vivo* and not *in vitro*, are to be found in GILLIES, D., *Philosophy of Science in the Twentieth Century. Four Central Themes*, Blackwell, Oxford, and Cambridge, MA, 1993, pp. 48-53.

type of penicillium, and most penicillia do not produce penicillin in the same quantity, if at all. If Fleming had ceased cultivating the mould, it would have been difficult to restart doing so. He continued the cultivation because he had found another use for mould juice.

The main source of income of the inoculation department where Fleming worked was the production and sale of vaccines. There was indeed an efficient unit for producing vaccine (a vaccine *laboratory*, as it was then called) within the walls of the department, and Fleming had been in charge of the production of vaccines since 1920. In particular, a vaccine was made against Pfeiffer's bacillus (*bacillus influenzae*) which was believed to cause influenza and other respiratory infections. It was difficulty to isolate this bacillus because cultures were apt to be swamped by other micro-organisms. Fleming, however, had discovered that penicillin, despite its effect on so many virulent bacteria, left Pfeiffer's bacillus unaffected. By incorporating penicillin into the medium on which he was growing Pfeiffer's bacillus, he could eliminate the other germs, and produce good samples of the bacillus itself. Fleming in fact used this method for preparing the influenza vaccine in his vaccine laboratory for this purpose every week after its discovery. Significantly, the title of Fleming's 1929 paper on penicillin was: "On the antibacterial action of cultures of a penicillium with special reference to their use in the isolation of *B. influenzae*." Because of this application of penicillin, cultures of the mould were established at the Lister Institute, Sheffield University Medical School, and at George Dreyer's School of Pathology at Oxford. Thus, when Florey and his team decided to take up again the question of whether penicillin might be a "perfect antiseptic," they were able to find samples of Fleming's strain of *penicillium notatum* just down the corridor in the Dreyer School of Pathology where they were working. This then is an appropriate moment to turn from Fleming to Florey and the Oxford team.

4. The Discovery of Penicillin Phase 2: The Work of Florey and his Oxford Team

As we have seen, Fleming worked for a while on lysozyme, and this led on to his later work on penicillin. Curiously enough the head of the Oxford team, the Australian Howard Florey, followed the same route. In section 4.1 it will be explained why Florey got interested in lysozyme, what research he and his team carried out on it, and why this research suggested that it might be useful to move on to investigate penicillin. Then in section 4.2 I will describe the Oxford team's work on penicillin between 6 September 1939 and 27 March 1943. It was this work which established that penicillin was, after all, a very effective antibiotic.

4.1. The Oxford Team also started with Lysozyme

Curiously enough the Oxford team also started by working on lysozyme, and moved on from there to penicillin. Howard Florey first got interested in lysozyme from his studies on mucus and its function. Lysozyme is found mainly in mucus-containing body fluids. At any rate on 17 January 1929 he sent various organs and tissues from experimental rats which had been killed to Fleming, presumably for help in assaying the quantity of lysozyme they contained.

In the late 1930s, Florey was joined in Oxford by Ernst Chain, a Jewish refugee from Nazi Germany, and an expert biochemist. During the academic year 1938-1939, Chain worked on lysozyme with Epstein, an American D. Phil. student and Rhodes scholar. They confirmed that lysozyme is an enzyme —a polysaccharidase, and then looked for the substrate in the bacterial cell wall which it broke down. This turned out to be N-acetyl glucosamine. This result was published in 1940.

While working on lysozyme, Chain surveyed the literature on other natural antibacterial substances which might be worth investigating. This is how he came across Fleming's 1929 paper on penicillin. This was in Vol. 10 of the *British Journal of Experimental Pathology*, while Fleming's papers on lysozyme were in Vols 3 and 8, and Florey's in Vol. 11. Chain at first thought that penicillin was a kind of mould lysozyme, a bacteriolytic enzyme on which he could repeat his investigation of lysozyme. Interestingly he did not realize at the time that Fleming was still alive.

After discussions with Florey, the two of them decided to investigate penicillin, and Florey prepared a grant application to support Chain.

4.2. The Oxford Team's Work on Penicillin (6 September 1939-27 March 1943)

Britain declared war on Germany on 3 September 1939, and on 6 September 1939 Florey sent in the grant application. Here is an extract:

> "Filtrates of certain strains of penicillium contain a bactericidal substance, called penicillin by its discoverer Fleming, which is especially effective against staphylococci, and acts also on pneumococci and streptococci. There exists no really effective substance acting against staphylococci *in vivo*, and the properties of penicillin which are similar to those of lysozyme hold out promise of its finding a practical application in the treatment of staphylococcal infections. Penicillin can easily be prepared in large amounts and is non-toxic to animals, even in large doses. Hitherto the work on penicillin has been carried out with very crude preparations and no attempt has been made to purify it. In our opinion the purification of penicillin can be carried out easily and rapidly.
>
> In view of the possible great practical importance of the above mentioned bactericidal agents it is proposed to prepare these substances in a purified form suitable for intravenous injections and to study their antiseptic action *in vivo*."[18]

In October the Medical Research Council approved Chain's grant for £300 per annum, with £100 expenses for three years. This was not enough but Florey got an additional grant from the Rockefeller Foundation of £1000 for the initial cost of equipment and £1670 per annum for 5 years. With this money, the work was able to go ahead.

The first step was to purify penicillin which Florey had rather optimistically said in his research proposal "can be carried out easily and rapidly." In fact it was a difficult task. The first result was that penicillin is soluble in alcohol. This had been shown earlier by S. R. Craddock and F. Ridley —Fleming's collaborators. This was interesting in that is showed that penicillin was not a protein. Chain confirmed this by demonstrating that penicillin would pass through micropore filters that retained proteins. So penicillin was not an enzyme and had a relatively small molecule. This must have disappointed Chain, but it did show that penicillin might be injectable without producing the allergic reactions due to foreign proteins.

The next result was that penicillin could be extracted by ether if the mixture was made acidic. This too had been discovered by a Professor of Biochemistry at the London School

[18] Taken from Macfarlane, G., *Howard Florey. The Making of a Great Scientist*, p. 299.

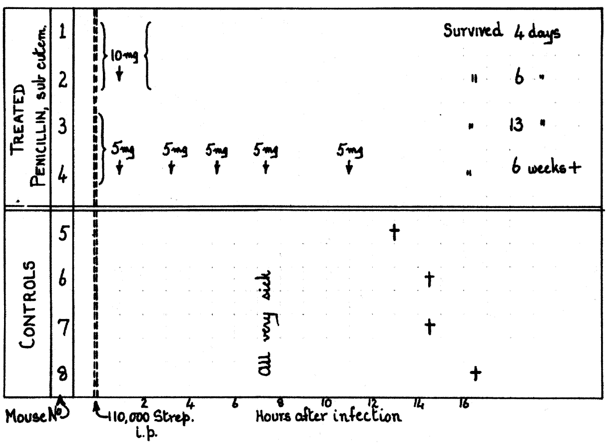
of Hygiene and Tropical medicine (Harold Raistrick) in 1930, but Raistrick had been unable to get the penicillin back from the ether. This problem was solved by another of Florey's team at Oxford: Norman Heatley.

In March 1940 Heatley discovered back-extraction. The idea was simple. If an acidified solution was needed to make penicillin dissolve in ether, perhaps an alkaline solution would cause it to come out of the ether. Heatley shook ether, containing penicillin, with alkaline water, and, sure enough, the penicillin passed back into the water. Curiously enough this had already been discovered by L. Holt working in Fleming's laboratory in 1934, but had never been published.

Partially purified penicillin could now be prepared, and Florey conducted a systematic series of administration and toxicity tests in April and May of 1940. Penicillin was injected into rats, mice, rabbits, and cats with no ill-effects, though it was rapidly excreted and had largely disappeared from the bloodstream in 1 or 2 hours. (Penicillin is in fact toxic to guinea pigs. So it is fortunate that these were not used.) Penicillin was also shown to be harmless to the delicate white cells known as leucocytes. Fleming's findings regarding the inhibitory effect of penicillin on a wide range of pathogenic bacteria were confirmed, and it was realised that penicillin worked by blocking the normal bacterial process of cell-division.

So far the Oxford team had largely repeated Fleming's work, albeit more systematically and on a larger scale. The next step was their crucial innovation. On Saturday 25 May 1940, they carried out the first mouse protection test. 8 mice were injected by Florey at 11am intraperitoneally with 100 million streptococci. 4 (the controls) received no further treatment. Of the remaining 4, 2 (Group A) received an injection of 10 mg of penicillin subcutaneously at 12 pm and no further treatment, while the remaining 2 (Group B) received 5 injections of 5 mg at 2 hour intervals starting at 12 pm. The results are shown in Figure 1.[19]

Figure 1. Chart showing the timing of injections and the results of the experiments drawn by Dr. N. G. Heatley from his notes of the first mouse protection test on 15 May 1940.



19 This figure is taken from MacFarlane, G., *Howard Florey. The Making of a Great Scientist*, p. 314.

The 4 control mice were all dead within 16 hours. The 2 group A mice survived 4 and 6 days. One group B mouse survived 13 days and the other 6 weeks +.

The result of this test was so striking that it could leave little doubt that penicillin was a highly effective antibiotic. However it was only the first of a series of animal trials carried out by the systematic Florey. These prepared the ground for the first clinical trial, but since a man is 3,000 times the weight of a mouse, the production of penicillin had to be greatly increased. Florey tried to interest several pharmaceutical firms but with little success. They were all occupied with seemingly more urgent war work. Undaunted, Florey decide to turn his laboratory into a factory, and was greatly helped in this by the innovative and resourceful Heatley. In 1941, the extraction of penicillin was improved by the use of column chromatography, which yielded penicillin ten times purer. This was necessary for the clinical trials, which were carried out by C. M. Fletcher, a doctor who worked with Florey.

The first clinical trial took place on 12 February 1941. This case illustrates the ravages which bacterial infections could cause in the pre-penicillin era. The patient was a 43 year old policeman, Albert Alexander, who had become infected after pricking himself on a rose bush. The infection had spread over his face, and his left eye had to be removed on 3 February. However, the infection spread further into his right shoulder and lungs. On 12 February, Fletcher gave him an intravenous injection of 200 mg of penicillin, followed by 100 mg at 3 hour intervals. The patient improved immediately. Within 4 days his temperature dropped to normal and his appetite returned. Unfortunately though, the penicillin was running out even though it was being extracted from his urine and re-injected. Without further penicillin, his improved health continued for a while, but then he relapsed and died on 15 March 1941.

The third patient, Percy Hawkins, a 42 year old labourer, was a more unequivocal success. He had a 4 inch carbuncle on his back. Beginning on 3 May 1941, he was given 200 mg of penicillin every hour for five hours, then 100 mg hourly. After 7 May this dose was reduced to 50 mg. By 10 May the carbuncle had almost completely disappeared. This was a striking case of a successful treatment of a localised staphylococcal infection.

Of course these clinical trials were only the start. Howard Florey continued the work helped by his wife Ethel who was a doctor. On 27 March 1943, they published a paper in *The Lancet* describing 187 cases of sepsis treated with penicillin. This established beyond doubt the efficacy of penicillin as an antibiotic.

5. Suggested Modifications of Kuhn's Theory in the Light of the Penicillin Example

Let us now return to Kuhn and see how well the work of Fleming and the Oxford team fits his theory of discovery in science. It is certainly a striking confirmation of Kuhn's claim: "… the discovery of scientific novelty … extends over time and may often involve a number of people."[20] And of Fleck's claim: "… discovery must be regarded as a *social event*."[21]

[20] Kuhn, Th. S., "The Historical Structure of Scientific Discovery," p. 171.

[21] Fleck, L., *Genesis and Development of a Scientific Fact*, p. 76.

This feature of scientific discovery is still largely unrecognised by the general public. Alexander Fleming is famous everywhere as *the* discoverer of penicillin, while the name of Howard Florey and the work of the Oxford team remain largely unknown.

This attribution of credit in the popular imagination goes back to the time (August 1942) when news first leaked out of the new "miracle cure" and the press got interested. A key factor was Sir Almroth Wright's letter to *The Times* of 31 August 1942 which ran:

> "Sir,
>
> In the leading article on penicillin in your issue yesterday you refrained from putting the laurel wreath for this discovery round anyone's brow. I would, with your permission, supplement your article by pointing out that, on the principle *palmam qui meruit ferat* it should be decreed to Professor Alexander Fleming of this laboratory. For he is the discoverer of penicillin and was the author of the original suggestion that this substance might prove to have important applications in medicine."[22]

In the outburst of press interest which followed, Fleming and Florey handled things very differently. Fleming saw reporters and gave interviews. Moreover this dour and laconic Scotsman proved to be a great favourite with the general public. Florey on the other hand refused to see any reporters and sent them away. It was natural then that the press should concentrate on Fleming whose name was the only one which became known to the public. The scientific community was more judicious. When a Nobel prize for the discovery of penicillin was awarded in 1945, it was divided between Fleming, Chain, and Florey. This was certainly quite reasonable, but it might well be argued that further members of the Oxford team, such as Norman Heatley, could have been included.

While the discovery of penicillin confirms Kuhn's views on the social nature of scientific discovery, it diverges from other aspects of his account. Kuhn essentially considers two types of discovery, viz. (i) *the unproblematic class*, in which an object is predicted and later discovered (e.g. radio waves), and (ii) *the troublesome class*, in which something strange is observed and its true nature only later elucidated. The discovery of penicillin does not fit exactly into either class. Penicillin was certainly not predicted, but, before its discovery, the concept of a perfect antiseptic had been formulated, and Fleming was on the look out for such an antiseptic without, however, any definite conviction that he would find one. It could be said that penicillin, or something similar, was hoped for rather than predicted.

In Kuhn's "troublesome class" of discoveries, more than one scientist may be needed because the process requires (i) the initial observation of something unusual (e.g. Herschel's observation of a curious celestial object), and (ii) the elucidation of the nature of what was observed (e.g. Lexell's claim that Herschel's celestial object was a planet). In the penicillin case, however, more than one scientist was necessary because Fleming's initial hypothesis regarding penicillin, viz. that it was a perfect antiseptic appeared to be refuted by experiments, so that further work was needed to show that the initial hypothesis was correct after all. Thus, in addition, to Kuhn's two classes, we might introduce a third class of discovery in which researchers are looking for something with a particular set of properties. The discovery then consists of two stages: (i) becoming aware of something which might have the required set of properties, and (ii) the demonstration that it really does have these properties.

---

[22] Macfarlane, G., *Howard Florey. The Making of a Great Scientist*, p. 349.

6. Bibliography

Fleck, L., *Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Einführung in die Lehre von Denkstill und Denkkollektiv*, Schwabe, Basel, 1935. Translation by Fred Bradley and Thaddeus J. Trenn: *Genesis and Development of a Scientific Fact*, The University of Chicago Press, Chicago and London, 1979.

Fleming, A., "On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of *B. Influenzae*," *British Journal of Experimental Pathology*, v. 10, (1929), pp. 226-236.

Gillies, D. A., *Philosophy of Science in the Twentieth Century. Four Central Themes*, Blackwell, Oxford and Cambridge, MA, 1993.

Hare, R., *The Birth of Penicillin and the Disarming of Microbes*, George Allen and Unwin, London, 1970.

Kuhn, Th. S., "The Historical Structure of Scientific Discovery," *Science*, v. 136, (1962), pp. 760-764. Reprinted in Kuhn, Th. S., *The Essential Tension*. *Selected Studies in Scientific Tradition and Change*, The University of Chicago Press, Chicago and London, 1977, Chapter 7, pp. 165-177.

Kuhn, Th. S., *The Structure of Scientific Revolutions*, The University of Chicago Press, Chicago and London, 1962.

MacFarlane, G., *Howard Florey. The Making of a Great Scientist*, Oxford University Press, Oxford, 1979.

MacFarlane, G., *Alexander Fleming. The Man and the Myth*, Chatto and Windus, The Hogarth Press, 1984.

Why Randomize? Evidence and Ethics in Clinical Trials

John Worrall

1. Introduction

In a randomized controlled experiment (henceforward an RCT) designed to test some new treatment —perhaps a new drug therapy or a new fertiliser— some experimental population (a set of patients suffering from some medical condition and recruited into the trial or a set of plots of land on which some crop is to be grown) is divided by a random process into two exhaustive and mutually-exclusive subsets: the "experimental group" and the "control group". Those in the experimental group receive the new test treatment while those in the control group do not. What they receive instead may differ from case to case: in agricultural trials, the control plots may simply be left unfertilised (all other obvious factors —such as frequency and extent of artificial watering, if any— being left the same as in the experimental regime), or, in clinical trials, the control group may be given either a "placebo" (generally a substance "known" to have no specific biochemical effect on the condition at issue) or the currently standard therapy for that condition. RCTs are almost invariably performed "double blind": neither the subjects themselves nor those treating them know whether they are receiving the experimental or the control treatment.[1]

It is widely believed that RCTs carry special scientific weight —often indeed that they are essential for any truly scientific conclusion to be drawn from trial data about the effectiveness or otherwise of proposed new therapies or treatments. This is especially true in the case of clinical trials (on which this paper will exclusively focus): the medical profession has been overwhelmingly convinced that RCTs represent the "gold standard" by providing the only truly "valid," unalloyed scientific evidence of the effectiveness of any therapy. Thus the statistician Tukey writes: "the only source of reliable evidence … is that obtained from … carefully conducted randomized trials," while Sheila Gore avers "Randomized trials remain the reliable method for making specific comparisons between treatments."[2] Clinical science may occasionally have to rest content (perhaps for practical reasons) with evidence from other types of trial, but this is always very much (at best) a case of (epistemic) second-best. For, it is widely believed, all non-randomized trials are inevitably subject to bias, while RCTs, on the contrary, are free from bias (or perhaps, and more plausibly, are as free from bias as any trial could possibly be).

In this paper I analyse the claim that RCTs provide far and away the best, perhaps the only really valid, scientific evidence; and I will closely scrutinise the arguments that have been put forward for it (section 1). In section 2 I shall consider a famous trial (or really a series of

[1] Other more complicated (and arguably more effective) randomized designs are possible —for example randomized blocks (where the two groups are matched for known prognostic factors and then which becomes the experimental and which becomes the control group is decided randomly). But the above is the simplest RCT and what most commentators have in mind in assessing the power of the methodology.

[2] Tukey, J. W., "Some Thoughts on Clinical Trials, especially Problems of Multiplicity," *Science,* v. 198, (1977), p. 1958; and Gore, S. M., "Assessing Clinical Trials-Why Randomize?," *British Medical Journal,* v. 282, (1981), p. 684.

trials) that shows that decisions have been made about which trials are ethical to perform that depend entirely on this epistemological claim that RCTs provide much stronger evidence than any other trial (and in particular than historically-controlled trials or "observational studies"). I will conclude (section 3) that because the arguments for the superior epistemic power of randomization seem on analysis to be weaker than they are widely considered to be, the trials outlined in section 2 —and no doubt others like them— raise enormous ethical concerns. And I also in that final section outline a seemingly more defensible approach to the question of the weight of evidence from clinical trials.

2. The Alleged Epistemic Superiority of RCTs

Many in the medical profession just accept it as a given that randomized controls are the proper scientific way of proceeding when it comes to judging the effectiveness of a new treatment. But it is not, of course, a given —instead it must be argued for on the basis of more fundamental principles of correct scientific method. The first question to ask from this more basic perspective is "Why "control" at all?".

The answer is clearly that we want to have evidence that any positive effect that we measure in a trial is genuinely attributable to the treatment under investigation. Suppose, to take a hoary old example, we are interested in the effect of giving patients suffering from common colds regular doses of vitamin C. We take a bunch of patients with colds, give them vitamin C, and then record, say, how many of the patients recover within a week. Suppose they all recover. It would nonetheless be an obvious mistake to infer from this that vitamin C is an effective treatment for colds. The first reason is the possibility that all those patients would have recovered from their colds within a week even had they not been given the vitamin C. If that were true, then to infer that vitamin C is an effective treatment would be to commit a particularly egregious version of the famous "post hoc ergo propter hoc" fallacy.

It would be ideal from the epistemological point of view if we could know what would have happened to those particular patients had they not been given the vitamin C; but of course we have no access to these counterfactual histories —all we know is that they were given the vitamin C and did recover within a week. The best we can do is take a different bunch of patients, also suffering from colds, and treat them differently —say in this case by doing nothing to them (in the jargon of the trade they would then form a "natural history control group").

Suppose that none of the patients in this "control group" recovers from his or her cold within the week. Given that those in the "experimental group" who were given vitamin C did recover, it would be tempting to infer that this (now) controlled trial had shown that vitamin C is effective. But a moment's thought shows that this too is premature. Suppose that those in the control ("natural history") group were all suffering from much heavier colds than those in the experimental group, or suppose that all those in the control group were older people suffering from other conditions as well whereas all those in the experimental group were, aside from their colds, young, fit and healthy. We intuitively want the two groups to be "equal (at least on average) in all other (relevant) regards." Only if we have evidence that all other factors were equal would we seem to be justified in claiming that the only difference between the two groups is the difference in treatment and hence be justified in claiming that any measured improvement in the experimental group relative to the control is caused by the treatment.

Now we can of course —at least in principle— ensure equality in the case of any factor that commonsense (or "background knowledge") tells us might well be relevant. This would be achieved by deliberately matching the two groups relative to the factor at issue. So, continuing with our hoary example, given that the severity of the cold is clearly relevant, we could match the two groups in this respect: either by ensuring that everyone in the trial had colds of (at least approximately) the same severity or by ensuring that the proportions of those with severe as opposed to mild colds is the same in both groups. Similarly since age and general health and fitness seem likely to be factors possibly relevant to early recovery, we can match the two groups so as to ensure that these factors are similarly distributed within the experimental and control groups.

Suppose we have matched the two groups with respect to every factor that anyone can think of as plausibly relevant to recovery. And suppose we find a substantially higher proportion of recoverers within the group given vitamin C. Surely now we have telling evidence that vitamin C aids recovery from colds?

Well, still "maybe not" and this for two separate reasons. First it may be (though we can't of course know it —a fact that we shall need to reflect on at length later) that the two groups now are indeed "equal" with respect to all factors relevant to recovery from colds (though it would actually be something of a miracle). In that case, we would indeed be justified in concluding that the observed difference in outcome was produced by the treatment rather than being due to some difference between the two groups. But it wouldn't follow that it was the fact that the treatment involved giving vitamin C that caused the improved outcome. It is conceivable —and there is a lot of evidence (albeit disputed by some commentators) that it is in fact the case— that a patient's expectations, fired by being treated by an authority figure, play a role in recovery from at any rate relatively minor complaints. In the vitamin C case, as we are now envisaging it, those in the control group receive no treatment at all —they are just the "otherwise equal in all respects" comparators. But the very fact that they receive no treatment —any treatment— at all is a difference and one that might, as just remarked, be relevant.[3]

This is the reason why the control groups in medical trials invariably will be either placebo controls or conventional treatment controls (sometimes there are two control groups in a trial: one of each). That is, those in the control group will not be left untreated, but will instead be treated either with the currently accepted treatment for the condition at issue or with a substance known to have no specific biochemical effect on the condition but which is intended to be indistinguishable to the patient from the (allegedly) "active" drug under test. This feature of clinical trials permits a further feature with clear methodological merits —namely that the trials can, again at least in principle, be performed "double blind". The two treatments ("active" drug and placebo (or conventional treatment)) can be sorted into packets marked simply, say, "A" and "B" by someone not involved in seeing patients and delivered to the patient in ways that are indistinguishable:

3 Notice though that the plausibility of this suggestion depends upon the particular case. If, for example, the patients involved are neonates, as in the case examined in section 2, then the idea that expectations may play a role in reaction to treatment can surely be dismissed.

hence neither the patient herself nor the clinician involved knows whether that particular patient has received the "active" drug or not.[4]

But even after matching with respect to factors that background knowledge says may plausibly play a role and even after using placebo or conventional treatment controls to try to ensure that no other difference is induced in what otherwise might have been "equal" groups just by the way the two groups are treated in the trial, there is still a further problem. This is that the list of factors that might make a difference to treatment outcome is of course endless. The groups may have been deliberately matched with respect to obvious factors such as severity of symptoms, sex, age, general level of health, and so on, but what if recovery from colds depends significantly on whether you were breast- or bottle-fed as a child, or on whether you have previously been infected with, and recovered from, some particular non-cold virus, or…? This is the problem of "unknown (better: unsuspected) factors" —by definition the groups cannot be matched with respect to unknown factors, so it is certainly possible that even groups perfectly deliberately matched on known factors are significantly different in respect of some unknown one, and therefore possible that the inference from an improved outcome in the experimental group to the effectiveness of the treatment (vitamin C in our hoary example) is invalid.

As we shall see, the strongest argument for the epistemic superiority of randomized trials (strongest in the sociological sense that it is the one that has convinced most people in medicine) is precisely that randomization is alleged to solve the problem of "unknown factors": a randomized trial is controlled for all factors known and unknown.

There are various versions of RCTs. In the most straightforward, the division of those involved in the trial into experimental and control groups is effected by some random process —individual patients being assigned to treatment "A" or treatment "B" according to the toss of a fair coin (or, in more approved manner, on the basis of a table of random numbers). In what seem to me (for reasons to be discussed) much more sensible trials —some stratification or "blocking" is carried out with respect to factors that might plausibly play a role in recovery (or whatever the outcome at issue in the trial is). Again the most straightforward version of this type of design (in principle, if not in practice) would be to take the whole population to be involved in the trial and divide it consciously into two groups that are matched with respect to "known" prognostic factors and then use a random process to decide which of these groups is given treatment A and which B.

So these are the main ideas behind RCTs. Why should such trials be thought to carry more epistemological weight than any other, non-randomized trial? There are, so far as I can tell, five different arguments to be found in the literature that claim to show why this is so.

---

4 This is not as straightforward as it sounds. The active drug will invariably have noticeable side-effects while the traditional bread- or sugar-pill will have none (or at least none worth the name in most cases: you wouldn't want to be too liberal with your sugar pills if diabetics were involved in your trials, nor with bread-pills if it involved some patients with gluten-intolerance). It would in such cases be easy for the blind to be 'broken' certainly for the clinicians involved and also for the subjects themselves. In recognition of this problem there are now attempts to perform trials involving 'active placebos' —substances that again are 'known' to have no characteristic effect on the condition but which mimic the side-effects of the (allegedly) active treatment under trial (See MONCRIEFF, J. ET AL., "Active placebos versus antidepressants for depression," *The Cochrane Database of Systematic Reviews* 2004, Issue 1, Art. No.: CD003012. DOI: 10.1002/14651858.CD003012.pub2.)

The first of these is Fisher's argument that randomization is necessary to underpin the logic of his famous significance test approach —universally applied (with some optional refinements) in medical trials. [5]

The second argument is the one we already noted as sociologically speaking the most persuasive —that by randomizing the trial is controlled not just for known factors but for all factors, known and unknown.

A third argument is based on the idea that standard methods of randomizing control, not for some hitherto unconsidered possible bias, but for a "known" bias that is believed to have operated to invalidate a number of trials —namely, "selection bias." If clinicians involved in trials are allowed to decide the arm of the trial that a particular patient is assigned to then there is the possibility that, perhaps subconsciously, they will make decisions that distort the result of the trial and thus produce an inaccurate view of the effectiveness of the treatment. They might, for example, have an opinion on the effectiveness of the new drug and also likely side effects, and therefore direct patients that they know to one arm or the other because of the perfectly proper desire to do their best for each individual patient, or because of the entirely questionable desire to achieve a positive result so as to further their careers or please their (often pharmaceutical company) paymasters. (I stress the "might" here.)

A fourth argument has also been given a good deal of emphasis of late within the Evidence-Based Medicine movement. This claims that, whatever the finer rights and wrongs of the epistemological issues, it is just a matter of historical fact that the "track-record" of RCTs is better than that of observational studies ("historically-controlled trials") because the latter standardly give unduly optimistic estimates of treatment effects.[6]

Fifthly and finally, an argument that randomization has special epistemic virtues has arisen from the currently burgeoning literature on "probabilistic causality". Several authors —notably Judea Pearl, David Papineau and Nancy Cartwright[7]— have argued that randomisation plays an essential role when we are seeking to draw genuinely causal conclusions about the efficacy of some treatment as opposed to merely establishing that treatment and positive outcome are associated or correlated.

Here I will concentrate on the second —and, as noted, most influential— argument, developing my analysis of it much further than in my earlier work.[8] And then I will briefly re-outline the analyses of the other four arguments that I have given in more detail elsewhere.

Mike Clarke, the Director of the Cochrane Centre in the UK, writes on their web-site: "In a randomized trial, the only difference between the two groups being compared is that of most interest: the intervention under investigation."[9]

---

5 Fisher's argument can be found in his book *The Design of Experiments,* Oliver and Boyd, London, 1935.

6 Historically-controlled trials (aka observational studies) are, as we shall see in more detail later, ones where the control group is considered to be supplied by (comparable) patients treated earlier with the previously accepted treatment. Hence in these trials, all patients actively involved are given the new treatment under test.

7 See PEARL, J., *Causality-Models, Reasoning and Inference,* Cambridge University Press, New York and Cambridge, 2000, PAPINEAU, D., "The Virtues of Randomization," *The British Journal for the Philosophy of Science,* v. 45, n. 2, (1994), pp. 437-450, and CARTWRIGHT, N., *Nature's Capacities and their Measurement,* Oxford University Press, Oxford, 1989.

8 I have analysed the first four of these arguments in WORRALL, J., "What Evidence in Evidence-Based Medicine?," *Philosophy of Science,* v. 69, n. S3, (2002), pp. S316-330, and the fifth in WORRALL, J. "Why There's no Cause to Randomize", *The British Journal for the Philosophy of Science,* forthcoming.

9 See the website of the Cochrane Collaboration at *www.cochrane.org.*

That is, by randomizing, all other factors —both known and unknown— are (allegedly) equalised between the experimental and control groups; hence the only remaining difference is exactly that one group has been given the treatment under test, while the other has been given, say, a placebo; and hence any observed difference in outcome between the two groups in a randomized trial (but only in a randomized trial) can legitimately be inferred to be the effect of the treatment under test.

Mike Clarke's claim is admirably clear and sharp, but it is clearly unsustainable (as indeed he himself later implicitly allows). Clearly the claim as made is quite trivially false: the experimental group contains Mrs Brown and not Mr Smith, whereas the control group contains Mr Smith and not Mrs Brown, etc. An apparently more plausible interpretation would take it as stating that the two groups have the same means and distributions of all the [causally?] relevant factors. It is not clear to me that this claim even makes sense, though, and, even with respect to a given (finite) list of potentially relevant factors, no one can really believe that it automatically holds in the case of any particular randomized division of the subjects involved in the study. Although many commentators often seem to make the claim (and although many medical investigators blindly following the "approved' methodology may believe it) no one seriously thinking about the issues can hold that randomization is a sufficient condition for there to be no difference between the two groups that may turn out to be relevant.

Here is an amusing example that illustrates this point. A study by L. Leibovici and colleagues was published in the *British Medical Journal* in 2001 entitled "Effects of remote, retroactive, intercessory prayer on outcomes in patients with bloodstream infection: randomised controlled trial."[10] The study looked at 3393 inpatients at the Rabin Medical Centre in Israel during 1990-1996 who had suffered from various bloodstream infections. In July 2000 (so, notice, between 4 and 10 years after they had suffered these infections), a random number generator was used to divide these patients into two groups; and which of the two became the treatment group was decided by a coin toss. 1691 patients were, so it turned out, randomized to the intervention group and 1702 to the control group. A careful check was made for "baseline imbalances" with regard to main risk factors for death and severity of illness. ("Baseline imbalances" are differences between the two groups in respect of known prognostic factors produced in a "purely" randomized trial —that is, one in which no deliberate attempt is made to make the two groups equal with respect to these known factors; and spotted after the random division has of course been made.) But no significant baseline imbalances were found. The names of those in the intervention group were then presented to a person "who said a short prayer for the well being and full recovery of the group as a whole.' Then, but only then, were the medical records of all the patients checked for those patients' mortality, for length of stay in hospital and for duration of the fevers they had suffered. The results were that mortality was 28.1% in the "intervention" group and 30.2% in the control group, a difference that orthodox statistical methodology declares "non-significant'; however both length of stay in hospital and duration of fever were significantly shorter in the

[10] Leibovici, L. et al., "Effects of remote, retroactive, intercessory prayer on outcomes in patients with bloodstream infection: randomised controlled trial," *British Medical Journal,* v. 323, n. 7327, (2001), pp. 1450-1451.

intervention group (p = 0.01 and p = 0.04).[11] Leibovici and colleagues drew the conclusion that "remote, retroactive intercessory prayer said for a group is associated with a shorter stay in hospital and shorter duration of fever in patients with bloodstream infection and should be considered for use in clinical practice."[12]

Although it ought to have been clear that the authors were writing with tongues firmly in cheeks (for example they remark that "no patients were lost to follow-up'!), the paper produced a heated discussion in the course of which some commentators seemed at least to be ready to take the result seriously. But even the most religiously-minded are surely unlikely to believe that the mysterious ways in which god sometimes moves include predicting at time t that some prayer will be said on behalf of some patients between 4 and 10 years later than t and intervening in the course of nature at t, on the basis of that prediction, to give those patients a better (overall) outcome!

Leibovici himself fully agreed with this as he made clear in the subsequent discussion: "If the pre-trial probability [of the eventual 'result'] is infinitesimally low, the results of the trial will not really change it, and the trial should not be performed. This, to my mind, turns the article into a non-study, though the details provided (randomization done only once, statement of a prayer, analysis, etc) are correct."[13] The sentiment, entirely in line with Bayesian, as opposed to classical statistical, methodology, is that we need to take into account not only the "improbability" of a particular outcome occurring if some "null hypothesis" is correct (here the null hypothesis is that there is no difference between the two groups despite the remote intercessory prayer "intervention" and that any observed difference in outcome between the two groups is due to chance), but also the prior probability of the "negation" of the null (here that the prayer really did have a retroactive effect).

But although Leibovici may not have intended the study to be taken seriously as a basis for "treatment," it *is* to be taken seriously as a criticism of orthodox statistical methodology and in particular of the suggestion that a properly randomized study always produces real evidence of effectiveness. Leibovici insisted, note, that "the details provided (randomization done only once, statement of a prayer, analysis, etc) are correct." So the fact is that this was a properly randomized study (in fact an exceptionally and impressively large one) that happened to produce what we take ourselves to know must be the "wrong" result. Obviously what must have happened here is that although the division into experimental and control groups was done impeccably and although the double blinding was equally impeccable(!), "by chance" some unknown confounder/s were unbalanced and produced the difference in outcome. Not only is this not impossible, it ought to happen, according to orthodox statistical methodology, on average in one in every 20 or so such trials on treatments that in fact have no effect (assuming that the standard 5% "significance level" is invariably used)!

---

11 These 'p values' mean that there was only a 1% chance of observing such a large difference in length of stay in hospital (or a still larger one) if the 'null hypothesis' (of exactly the same probability in the two groups of staying in hospital for any given period) were correct; and only a 4% chance of observing such a large difference in duration of fever (or a still larger one) if the corresponding null hypothesis were correct.

12 Leibovici, L. et al., "Effects of remote, retroactive, intercessory prayer on outcomes in patients with bloodstream infection: randomised controlled trial," p. 1451.

13 Leibovici, L., "Author's Comments," *British Medical Journal,* v. 324, (2002), p. 1037.

The fact that —despite assertions like the one quoted from Mike Clarke— a particular random division may of course produce an importantly unbalanced division is indeed implicitly admitted even by the most orthodox advocates of randomization when they accept that if a particular ("pure") randomisation (involving no element of prior deliberate matching) has produced an imbalance in a "known" prognostic factor then one should not proceed to make any inferences. Though —surely rather quixotically— the orthodox then go on to assert that, in this situation, rather than deliberately match "known" factors, one should re-randomise until a random division is produced about which one has no concerns from the point of view of imbalances.

In sum, despite what is sometimes written, no one can seriously believe that having randomized is a sufficient condition for a trial result to reflect the true effect of some treatment. Is randomizing a necessary condition for this? That is, is it true that we cannot have real evidence that a treatment is genuinely effective unless it has shown itself to be so in a properly randomized trial? Again, some people in medicine sometimes talk as if this were the case, but again no one can seriously believe it. Indeed modern medicine would be in a terrible state if it were true. The overwhelming majority of all treatments regarded as unambiguously effective by modern medicine today —from aspirin for mild headache through diuretics in heart failure and on to pretty well any surgical procedure (appendicectomy, cholecystectomy, etc., etc.)— were never (and now, let us hope, never will be) validated in an RCT. Much of the impetus behind the "Evidence-Based Medicine" movement that emerged in the 1980s was the recognition that certain treatments that had been in regular use in medicine (such as grommets for glue ear, suppression of ventricular ectopic beats in heart failure and a *few* others) proved, when subject to systematic testing, in fact to be either ineffective or (worse) dangerous because adverse side-effects overwhelmed any positive effect on the target disorder. Although this is true, we must not let it blind us to the fact that the overwhelming majority of treatments in medicine that no one suggests for a moment are ineffective have never been subjected to an RCT.[14]

The above criticism —particularly of the alleged sufficiency of randomisation to establish that a treatment is effective— will be regarded by some as an attack on a strawman. Maybe this strawman produces real writing but if so it is of self-consciously simplified accounts aimed at medical practitioners (or perhaps those involved with the administration of research) with no knowledge of, or taste for, statistical niceties. The serious claim is, not that in a randomized trial all other factors aside from the treatment are automatically equal in the two groups, but rather that this is highly probable. A positive result in a randomized test, because the two groups are probably equal in all other respects, gives us not of course foolproof, but still the best evidence of treatment effectiveness that we could possibly have. We do not eliminate entirely the possibility of "bias" by randomizing, but we do "eliminate" it "in some probabilistic sense."

The problem is that for all its seeming plausibility and indeed for all its widespread acceptance and therefore immense practical significance, it seems difficult to make anything like full sense of this claim —especially on the basis of the orthodox approach to statistics. The latter (officially) refuses to deal in the probability of hypotheses at all,

[14] Nor should it blind us to the fact that it has also sometimes turned out to be true that treatments 'validated' in RCTs have later been withdrawn because of negative side-effects.

but only in the acceptance or rejection of hypotheses that attribute some probability to the values of some random variable. In order even to begin to make sense of the claim, we would need to be able to show that, for any particular (potentially) prognostic factor aside from whether a patient is given the treatment under test or, say, placebo, it is probable that that extra factor is identically (or near identically?) distributed in the two groups —treatment and control. Any plausibility that such a claim might appear to have depends, however, on confusing what can reasonably be asserted in the case of a single random division with what might reasonably be asserted about an indefinite number of repetitions of the random division.

What can it mean to claim that it is improbable that factor X is evenly distributed between the two groups? Assuming the classical non-Bayesian approach to probability (and there is no direct role for randomization according to the Bayesian approach),[15] it can only be a claim about an indefinite series of repetitions of the trial: that if you were to take a population and divide it at random into two lots and lots of times and record the cumulative relative frequency of positive values of X in the two groups (assume for simplicity that X is a two-valued random variable), then in the indefinite long run that frequency would be the same in the experimental and control groups and in fact the same as the actual frequency of positive values of X in the study population as a whole. But medical researchers involved in some particular trial do not make a random division indefinitely often, they do it once! In that one trial, factor X may be as substantially unbalanced between the two groups as you like, and there seems just to be no way to quantify what the "probability" of a substantial imbalance is: "single case probabilities" not being properly defined on this approach. Once you further take into account the fact that, by definition, the list of possible "unknown" factors is indefinitely long, then matters become even murkier. Even if one wanted to insist that despite the lack of any adequate formal analysis it was somehow "intuitively" clear that for any single factor X, it is "improbable" that it is significantly maldistributed between the two groups in a single randomisation, it would not of course follow even "intuitively" that it is improbable that there is no factor relative to which the single randomization is unbalanced —because of the lack of any real grasp of the list of potential other factors and of how large it is, this just seems to be, even intuitively, undefined.[16]

It is, then, difficult to see any objective weight in what is, as I mentioned earlier, sociologically speaking the most persuasive argument for the special epistemic power of randomized trials. What of the other arguments?

Let's start by accentuating the positive. There seems no doubt that the claim that randomization controls, not for all factors, but for the single "known" factor of "selection bias," carries some weight. This is accepted even by Bayesian critics of randomisation. If the experimenters are allowed consciously or unconsciously to influence the membership of the experimental and control groups, then this is undoubtedly a source of possible bias and hence an alternative (partial) explanation of any positive (or negative) result achieved. Notice however two things. First, this argument does not supply any reason to think that

---

[15] See for example KADANE, J. B. and SEIDENFELD, T., "Randomization in a Bayesian Perspective," *Journal of Statistical Planning and Inference,* v. 25, (1990), pp. 329-345.

[16] See for example LINDLEY, D. V., "The Role of Randomization in Inference," *PSA 1982,* volume 2, (1982), pp. 431-446.

the randomization itself has an effect —the procedure involved in randomized trials is just one way of ensuring that the clinicians are blinded to the group allocation of individual patients. If clinicians were prevented in some other way from influencing the allocation, or if it is clear for other reasons that they did not influence it, then randomization would, so far as this particular argument goes, become redundant. Secondly, as even Doll and Peto, the staunchest of RCT-advocates, allow,[17] selection bias, where operative (or possibly operative) at all, is unlikely to produce a large effect. It would therefore be a mistake to dismiss a non-randomized study that had revealed a large effect as not providing any "valid" evidence simply on the grounds that selection bias might have played a role.

Fisher's famous argument for randomization was that it is the only way in which the logic of his method of statistical significance testing can be underwritten. He argued in effect that when, but only when, you have randomized, is it legitimate to identify the "null hypothesis" (which would otherwise just be the very vague hypothesis that any positive result observed is not the result of the treatment being tested) with the "chance" hypothesis that each patient has the same probability of recovering (or whatever the outcome measure is) independently of whether they received the treatment under test (or the placebo/ conventional treatment). Fisher's argument, based on his "Tea Lady" test (perhaps the most celebrated in the whole history of statistics), has appeared compelling to many and it does have a tenacious air of plausibility. However Bayesians have in my view conclusively shown that it holds no water. Since the considerations raised are rather technical, I shall not pursue this matter here but just refer to what seems to me the clearest Bayesian demolition —that by Colin Howson.[18] There is, in any case, an increasingly popular view (one that I share) that the whole classical statistical significance test methodology is itself fundamentally illogical and should be abandoned. If this view is correct then it would of course follow that, even were Fisher right that randomization is necessary to underpin the application of his methodology, this would still supply no cogent reason to randomize.

An argument that has had some impact in the recent Evidence-Based Medicine movement claims that whatever the finer rights and wrongs of the epistemological issues it is just a matter of fact that the "track-record" of RCTs is better than that of other types of study, and in particular better than the track record of so-called historically controlled studies. In this latter type of study, the control group is supplied by an allegedly comparable set of previous patients treated in the (recent) past with whatever is currently accepted treatment. Hence in such studies (also sometimes called "observational studies") all the patients actively involved in the trial are given the new treatment under test. The claim is that these studies standardly give unduly optimistic estimates of treatment effects. This argument, so I suggest in my paper "What Evidence in Evidence-Based Medicine?," is (a) circular (it depends on supposing that, where an RCT and an "observational study" have been performed on the same treatment, it is the former that reveals the true efficacy (after all randomized results provide the "gold standard"!) and this is precisely the question at issue; (b) based largely at least on comparing RCTs to particularly poorly performed observational studies that anyone would agree are obviously methodologically unsound;

[17] See Doll, R. and Peto, R., "Randomized Controlled Trials and Retrospective Controls," *British Medical Journal,* v. 280, (1980), p. 44.

[18] See Howson, C., *Hume's Problem,* Oxford University Press, Oxford, 2000, pp. 48-51.

and (c) is —to say the least— brought into severe question by more recent work that seems to show that, where a number of different trials have been performed on the same treatment, the results of those done according to the RCT protocol differ from one another much more markedly than to do carefully performed and controlled observational studies.[19]

The final argument for randomization is one that has emerged from the recent literature on "probabilistic causality". A number of authors —taking routes that are related though different in details— have claimed that it is only when you randomize that a trial can give evidence of a genuinely causal (as opposed to merely "associational") connection between the treatment and outcome. A central problem in this area of probabilistic causality is that of distinguishing between "real" (causal) and "spurious" correlations. Two variables may covary despite being causally unconnected with one another —they might, for example, be two independent effects of a "common cause". So, to take an obvious example, the probability that you will develop lung cancer is much higher if you own a reasonable number of ashtrays (say more than 3) than if you don't:

P (lung cancer / own more than 3 ashtrays) >> P (lung cancer).

Thus lung cancer and ashtray ownership are (strongly) probabilistically associated (or "correlated" as is often said in this literature —though this is not the usual statistical meaning of the term.) But we wouldn't say that owning ashtrays "increases the probability" of developing lung cancer, because there is, as we know on the basis of background knowledge, no causal connection between the two. The causal connections are instead between smoking cigarettes and developing lung cancer, and smoking cigarettes and "needing" ashtrays. In the jargon, smoking cigarettes is a common cause of both lung cancer and ashtray-ownership. The fact that this is so and hence that the "correlation" between cancer and ashtrays is "spurious" is revealed by the fact that smoking "screens off" cancer from ashtray-ownership. In other words, the conditional dependence between the latter two variables disappears when you further conditionalise on smoking:

P (lung cancer/ own more than 3 ashtrays and you smoke) = P (lung cancer / you smoke),

even though P (lung cancer / own more than 3 ashtrays) >> P (lung cancer).

The argument then in essence (and ignoring some important issues about the inference from (observed) relative frequencies to (theoretical) population probabilities) is that you are justified in taking an observed relationship between treatment and whatever your outcome measure is (recovery within some fixed period, say) when, but only when, this relationship is observed in a trial that was randomized. In effect, then, the claim is that randomization eliminates the possibility of a "common cause" of treatment and treatment outcome.

In a forthcoming paper,[20] I take the various versions of this argument —by Nancy Cartwright, David Papineau, and especially Judea Pearl— and show that they fail. I shall not repeat the details of my counterargument here. But, as the above brief outline will perhaps suggest, their claim is at root just a particular version of the "randomizing controls for all other factors" line and hence it falls to the same objection: that it trades on a confusion between what might be justified in the indefinite long run of reiterated randomizations on

---

[19] For detailed references, see Worrall, J., "What Evidence in Evidence-Based Medicine?," *Philosophy of Science,* v. 69, n. S3, (2002), pp. S316-330.

[20] Cf. Worrall, J., "Why There's no Cause to Randomize," *The British Journal for the Philosophy of Science,* forthcoming.

the same group and what is justified in the particular case where, by definition, the random division has been effected only once. It is of course possible in the single case that the two groups are unbalanced in respect of a factor that is in fact a common cause of treatment and treatment outcome.

No argument known to me, then, really establishes the almost universally held view that RCTs have a special epistemic status —except for the modest argument about controlling for "selection bias"(and that bias might be eliminable by other means). In the next section, I will look at some trials on a newly-introduced treatment that were motivated entirely by the view that only evidence for treatment efficacy from an RCT really counts scientifically. The lack of any substantial and cogent argument for the necessity of randomization makes the ethical acceptability of these trials extremely suspect. In the final section, I will make some suggestions about what seem to me the correct ethical and methodological judgements.

3. Why the Issue is of Great Practical and Ethical Significance-The ECMO Case[21]

A persistent mortality rate of more than 80% had been observed historically in neonates experiencing a condition called persistent pulmonary hypertension (PPHN). A new method of treatment —using a technique developed for other conditions and called "extracorporeal membranous oxygenation" (ECMO)— was introduced in the late 1970s, and Bartlett and colleagues at Michigan found, over a period of some years, mortality rates of less than 20% in these infants treated by ECMO.[22] I think it is important background information here that this new treatment could hardly be regarded as a stab in the dark. It was already known that the underlying cause of this condition was immaturity of the lungs in an otherwise ordinarily developed baby. The babies that survived were those that somehow managed to stay alive while their lungs were developing. ECMO in effect takes over the function of the lungs in a simple and relatively non-invasive way. Blood is extracted from the body before it reaches the lungs, is artificially oxygenated outside the body, reheated to regular blood temperature and reinfused back into the baby —thus bypassing the lungs altogether.

Despite the appeal of the treatment and despite this very sharp increase in survival from 20% to 80% the ECMO researchers felt forced to perform an RCT ("... we were compelled to conduct a prospective randomised study") even though their experience had already given them a high degree of confidence in ECMO ("We anticipated that most ECMO patients would survive and most control patients would die...") They felt compelled to perform a trial because their claim that ECMO was significantly efficacious in treating PPHS would, they judged, carry little weight amongst their medical colleagues unless supported by a positive outcome in such a trial.[23] These researchers clearly believed that, in effect, the long established mortality rate of more than 80% on conventional treatment provided good enough controls that babies treated earlier at their own and other centres with conventional medical treatment

---

[21] It was Peter Urbach who first drew my attention to this case.

[22] See Bartlett, R. H., Andrews, A. F. et al., "Extracorporeal Membrane Oxygenation for Newborn Respiratory Failure. 45 Cases," *Surgery,* v. 92, n., (1982), pp. 425-433.

[23] This is another argument for RCTs that is not infrequently cited by medics and clinical scientists. It is however a very strange argument: if it were the case that randomizing was, in certain cases, neither necessary nor useful then it would seem better to try to convince the medical profession of this rather than turn their delusions into an argument for pandering to that delusion!

provided sufficiently rigorous controls; and hence that the results of around 80% survival that they had achieved with ECMO already showed that ECMO was a genuinely efficacious treatment for this dire condition. Given that there was an argument for thinking that there was no significant difference between the babies that Bartlett and colleagues had been treating using the earlier techniques and those that they had now been treating with ECMO (we will return to this point later), this counts as a (retrospective) historically controlled trial —one producing a very large positive result. But, because historically controlled trials are generally considered to carry little or no weight compared to RCTs, as we saw in the previous section, these researchers felt forced to go ahead and conduct the trial.

They reported its outcome in 1985[24]. Babies suffering from PPHN were allocated to ECMO treatment or to the control group (receiving the then conventional medical therapy —CT) using a modified protocol called "randomised play the winner". This protocol involves assigning the first baby to treatment group purely at random —say by selecting a ball from an urn which contains one red (ECMO) and one white (CT) ball; if the randomly selected treatment is a success (here: if the baby survives), then an extra ball corresponding to that treatment is put in the urn, if it fails then an extra ball corresponding to the alternative treatment is added. The fact that this protocol, rather than pure randomization, was used was no doubt itself a compromise between what the researchers saw as the needs of a scientifically (or is it sociologically?) convincing trial and their own convictions about the benefits of ECMO.

As it turned out, the first baby in the trial was randomly assigned ECMO and survived, the second was assigned CT and died. This of course produced a biased urn, which became increasingly biased as the next 8 babies all happened to be assigned ECMO and all turned out to survive. The protocol, decided in advance, declared ECMO the winner at this point, though a further two babies were treated with ECMO (officially "outside the trial") and survived. So the 1985 study reported a total of 12 patients, 11 assigned to ECMO all of whom lived and 1 assigned to CT who died. (Recall that this is against the background of a historical mortality rate for the disease of around 80%.)

Ethics and methodology are fully intertwined here. How the ethics of undertaking the trial in the first place are viewed will depend, amongst other things, on what is viewed as producing scientifically significant evidence of treatment efficacy: clearly a methodological/ epistemological issue. If it is assumed that the evidence from the "historical trial" (i.e. the comparison of the results using ECMO with the earlier results using CT) was already good enough to give a high degree of rational confidence that ECMO was better than CT, then the ethical conclusion might seem to follow that the death of the infant assigned CT in the Bartlett study was unjustified.

But if, on the other hand, it is taken that

> "... the only source of reliable evidence about the usefulness of almost any sort of therapy ... is that obtained from well-planned and carefully conducted randomized ... clinical trials,"[25]

---

[24] See Bartlett, R. H., Roloff, D. W., et al., "Extracorporeal Circulation in Neonatal Respiratory Failure: A Prospective Randomized Study," *Pediatrics,* v. 76, n., (1985), pp. 479-487.

[25] Cf. Tukey, J. W., "Some Thoughts on Clinical Trials, especially Problems of Multiplicity," *Science,* v. 198, (1977), p. 1958. (Emphasis supplied)

then you're likely to have a different ethical view, even perhaps that

> "the results [of the 1985 study] are not ... convincing... Because only one patient received the standard therapy, ..."[26]

Many commentators in fact took this latter view and concluded that

> "Further randomized clinical trials using concurrent controls and ... randomisation ... will be difficult but remain necessary."[27]

Those taking this second view held that neither the "historically controlled" results (i.e. the comparison of the mortality rates achieved with ECMO with the historical mortality rate achieved with conventional treatment) nor the results from this initial "randomized play the winner" trial had produced any reliable, scientifically-telling information. The Michigan trial had not produced any real evidence because —in deference to the researchers" prior views— it had not been "properly randomised". Indeed, they even imply (note their "will be difficult" remark) that such trials and their "historically controlled" antecedents, have, by encouraging the belief that a new treatment is effective in the absence of proper scientific validation, proved pernicious by making it more difficult to perform a "proper" RCT: both patients and more especially doctors find it harder subjectively to take the "objectively-dictated" line of complete agnosticism ahead of "proper" evidence. Some such commentators have therefore argued that historical and non-fully randomized trials should be actively discouraged. (Of course since historical trials in effect always happen when some new treatment is tried instead of some conventional treatment, this really amounts to the suggestion that no publicity should be given to a new treatment, and no claims made about its efficacy, ahead of subjecting it to an RCT.)

In the ECMO case, this line led to the recommendation of a further, and this time "properly randomized," trial which was duly performed. This second trial involved a fixed experimental scheme requiring $p < .05$ with conventional randomization but with a stopping-rule that specified that the trial was to end once 4 deaths had occurred in either experimental or control group. A total of 19 patients were, so it turned out, involved in this second study: 9 of whom were assigned to ECMO (all of whom survived) and 10 to CT (of whom 6 survived, that is 4 died). Since the stopping-rule now specified an end to the trial but various centres were still geared up to take trial-patients, a further 20 babies who arrived at the trial centres suffering from PPHS were then all assigned to ECMO (again officially "outside the trial proper") and of these 20 extra patients 19 survived.[28]

Once again, views about the ethics of this further trial and in particular about the 4 deaths in the CT group will depend on what epistemological view is taken about when it is or is not reasonable to see evidence as validating some claim. If it is held that the first trial was indeed methodologically flawed (because "improper" randomization had

---

[26] See Ware, J. H. and Epstein, M. D., "Comments on 'Extracorporeal circulation in neonatal respiratory failure: A prospective randomized study' by R. H. Bartlett et al.," *Pediatrics,* v. 76, (1985), pp. 849-851.

[27] Ware, J. H. and Epstein, M. D., "Comments on 'Extracorporeal circulation in neonatal respiratory failure: A prospective randomized study' by R. H. Bartlett et al.," p. 851.

[28] O'Rourke, J. P. et al., "Extracorporeal membrane oxygenation and conventional medical therapy in neonates with persistent pulmonary hypertension of the new born: A prospective randomized study," *Pediatrics,* v. 84, (1989), pp. 957-963.

resulted in only one patient being in the control group) and therefore that no real objective information could be gathered from it, then the conviction that the first trial result (let alone the historically controlled evidence) had already shown that ECMO was superior was merely a matter of subjective opinion. Hence this second trial was necessary to obtain proper scientific information.[29] On the other hand, if the correct methodological judgment is that the evidence both from previous practice and from the initial trial was already rationally compelling, then this second trial, and the deaths of 4 infants treated by CT in it, would seem to be clearly unethical.

Nor was this the end of the matter. Stopping rules of the sort employed in this second trial are anathema to orthodox statisticians, despite seeming entirely sensible from an intuitive point of view. (Surely another reason to be sceptical of that orthodoxy.)[30]

Hence many statisticians argued that even this second trial had not produced truly reliable scientific evidence of the effectiveness of ECMO for PPHN. Stuart Pocock, for example, wrote:

> "... a decision was taken to halt randomization when the data disclosed four deaths among ten infants receiving conventional medical treatment compared with none among nine infants having ECMO (p = 0.054) [R]andomization was stopped early on the basis of a fairly small amount of data, all subsequent patients being allocated to ECMO.
>
> The investigators were sensitive to the individual ethics of seeking parental consent and randomization for the next newborn infant ... However, with only 19 patients this does not represent strong evidence of the superiority of ECMO and provides little scope for making reliable judgments on the benefits of this treatment for universal use in such newborn infants in the future.
>
> Thus collective ethics may have been compromised by such early stopping…. [I]f ECMO really is effective the prolonged uncertainties maintained by lack of really substantial evidence may well have led to fewer newborn infants worldwide receiving it than would have been the case had the trial continued longer."[31]

But surely we would feel no such "ethical conflict" if we did not in fact already feel that we had sufficient evidence that ECMO is effective both from the "observational study" (that is, the striking comparison between the results of treating babies with PPHN using ECMO compared to what had historically been achieved using the earlier treatment) and from the earlier trials. Of course we need to distinguish carefully between merely subjective opinion and genuine evidence; and it is true that clinicians have occasionally

---

[29] There is still then of course the central, and in my view ultimately irresolvable ethical issue of the extent to which it is ethically acceptable to inconvenience (or worse) those patients involved in the trial for the promise of establishing a new treatment as effective for the benefit of future patients.

[30] For an account —and detailed criticism— of the reasons statisticians disapprove of stopping rules, see Howson, C. and Urbach, P., *Scientific Reasoning: the Bayesian Approach,* Second edition, Open Court, La Salle, IL, 1993, Chapter 9.

[31] See Pocock, S. J., "Statistical and Ethical Issues in Monitoring Clinical Trials," *Statistics in Medicine,* v. 12, (1993), pp. 1459-1469. For Pocock, "The basic ethical conflict in monitoring trial results is to balance the interests of patients within the trial —that is, the individual ethics of randomizing the next patient— and the longer term interest of obtaining reliable conclusions on sufficient data —that is, the collective ethics of making appropriate treatment policies for future patients," Pocock, S. J., "Statistical and Ethical Issues in Monitoring Clinical Trials," p. 1459.

firmly believed that they “knew” that certain treatments are effective that we now have seemingly-conclusive evidence are not; but, unless some reason can be given why RCTs should be so much superior in evidential value that nothing else really counts (and I have suggested that no such reason has so far been given), then we seem to be in a situation where proper scientific judgment is in conflict with the imposed statistical orthodoxy. The sensible solution would seem to be to reject the orthodoxy —there would then be no ethical conflict, but simply the “individual” ethical conviction that assigning any baby to conventional treatment in any of these trials was unethical!

One consequence of this view of Pocock’s and of other statisticians was yet another “properly randomised” trial on ECMO for the treatment of PPHN —this one in the UK. It was stopped early by the “oversight committee” because of an excess of deaths in the conventional treatment arm. (Oversight committees are independent bodies who play no role in the treatment or assignment of patients in the trial, but who are allowed to keep a running total of the outcomes on the two arms and intervene where this seems to them the ethical thing to do. Needless to say such committees, like stopping rules, are anathema to classical statistical orthodoxy.)

4. Conclusion: Towards a More Defensible Approach to Evidence in Clinical Trials

We saw in Section 2 that there is —at least— some reason to be suspicious of the cogency of all of the arguments for the superior epistemic power of RCTs, except for the modest one that randomizing controls for “selection bias”. Is it plausible that selection bias might invalidate the “observational” evidence for ECMO’s effectiveness? Well, clearly if Bartlett and colleagues were achieving an 80% survival rate by dint of carefully selecting only some privileged subset of the babies that presented at their hospital with PPHN —those with relatively mild versions of the condition, those whose lungs were closest to normal development or whatever— then we would have good reason to be sceptical that their results were genuine evidence for the effectiveness of ECMO: perhaps CT would have achieved 80% survival in the same highly selected group (or even better!). There has never been any suggestion, however, so far as I am aware, that this was the case. Bartlett and colleagues seem just to have begun to treat all the babies that arrived in their hospital with PPHN, and who would earlier have been given CT, with ECMO. There is also no serious suggestion that the demographics of the catchment area of the Michigan University Hospital changed in any significant way or that the nature of the condition changed in some way or any other reason why the group against which the ECMO treated babies were compared was in any significant way different from it.

But these are surely the questions that should have been asked rather than the insisting on an RCT. What is operating, at root, in the methodology of trials is something like Mill’s methods —the “controls” are in essence out to eliminate other potential explanations of any observed positive effect aside from the explanation that it is the effect of the treatment at issue. We cannot, I suggest, do any better than control for factors that background knowledge exhibits as plausible (partial) “causes” of response to treatment. I can see, as explained earlier, no basis for the belief that we can do better than this by randomizing —we are always at the mercy of the possibility that some other “unknown” factor happens to be significantly unbalanced between experimental and control groups and that it, rather than some specific effect of the treatment, is what explains (or perhaps chiefly explains) a positive outcome in a trial. It would be good to have grounds to eliminate that possibility but neither randomization nor anything else can supply them. We just do have to rely on plausibility assumptions grounded in "background knowledge."

This is not to say that randomization should always be avoided —it usually does no harm and may do some good. It does some good in terms of controlling for selection bias. Though I should again reiterate that, despite another well-entrenched myth, there seems to be no epistemological foundation for the view that randomization is the only way to eliminate selection bias: if Bartlett and colleagues, for example, simply treated every baby who would have been treated with the then conventional treatment with ECMO, then there could be no selection bias (and if they did not then there would be hospital records of PPHN babies being rejected for ECMO treatment). Randomization does no harm on two conditions. The first is that known prognostic factors are equally balanced —you surely do not want to leave it in the lap of the dice-playing gods whether the trials groups are balanced for obvious factors; I can see no epistemological virtue in the rigmarole of checking, after the event, for "baseline imbalances" rather than matching from the beginning, but so long as the groups do finish up reasonably well matched in terms such as age, general level of health and fitness and so on then it doesn't matter how exactly they get there. The second condition under which randomization does no harm is that some sort, any sort of trial, should be justified —this condition is not satisfied if there is evidence already available (as there arguably was in the ECMO case) that establishes the superiority of the new treatment.[32] (This will not in fact often be the case even if I am right that well-thought-through historically controlled trials are in principle just as telling as RCTs —the sort of very large effect produced by ECMO is unusual in current medicine and clearly the smaller the effect the more uncertain the evidence.[33]) We need the sort of systematic thought about plausible "confounders" that is suggested by Mill's methods and that forms the basic method of enquiry throughout epidemiology. The question to ask was not "has ECMO been 'validated' in an RCT?" but rather "Is there any plausible alternative explanation for such a large change in the mortality rate other than that it was produced by the change to treatment by ECMO?" It seems clear that the answer is "no" and that the evidence produced by the historical comparisons, because it was of the outcomes for so many more babies, in fact was weightier than that produced in either of the first two subsequent trials —"properly randomized" or not!

5. Bibliography

Bartlett, R. H., Andrews, A. F. et al., "Extracorporeal Membrane Oxygenation for Newborn Respiratory Failure. 45 Cases," *Surgery,* v. 92, (1982), pp. 425-433.

Bartlett, R. H., Roloff, D. W. et al., "Extracorporeal Circulation in Neonatal Respiratory Failure: A Prospective Randomized Study," *Pediatrics,* v. 76, (1985), pp. 479-487.

[32] Of course "establishes" here has the sense of "establishes defeasibly but as well as we possibly can at present." Equally obviously the neonates given ECMO should have been (and actually were) carefully monitored for possible side-effects of the treatment. But the side-effects issue gives no reason to prefer RCTs to historical trials.

[33] Not that there aren't troubling ethical issues about so-called mega-trials aimed at 'establishing' very small effects of treatments for very common conditions. See Penston, J., *Fiction and Fantasy in Medical Research: the Large-Scale Randomised Trial,* The London Press, London, 2003.

CARTWRIGHT, N., *Nature's Capacities and their Measurement,* Oxford University Press, Oxford, 1989.

DOLL, R. and PETO, R., "Randomized Controlled Trials and Retrospective Controls," *British Medical Journal,* v. 280, (1980), p. 44.

FISHER, R. A., *The Design of Experiments,* Oliver and Boyd, London, 1935.

GORE, S. M., "Assessing Clinical Trials-Why Randomize?," *British Medical Journal,* v. 282, (1981), pp. 679-684.

HOWSON, C., *Hume's Problem,* Oxford University Press, Oxford and New York, 2000.

KADANE, J. B. and SEIDENFELD, T., "Randomization in a Bayesian Perspective," *Journal of Statistical Planning and Inference,* v. 25, (1990), pp. 329-345.

LEIBOVICI, L., "Effects of Remote, Retroactive, Intercessory Prayer on Outcomes in Patients with Bloodstream Infection: Randomised Controlled Trial," *British Medical Journal,* v. 323, n. 7327, (2001), pp. 1450-1451.

LEIBOVICI, L., "Author's Comments," *British Medical Journal,* v. 324, (2002), p. 1037.

LINDLEY, D. V., "The Role of Randomization in Inference," *PSA 1982,* volume 2, (1982), pp. 431-446.

MONCRIEFF, J. ET AL., "Active Placebos versus Antidepressants for Depression," *The Cochrane Database of Systematic Reviews* 2004, Issue 1, Art. No.: CD003012. DOI: 10.1002/14651858.CD003012.pub2

O'ROURKE, J. P. ET AL., "Extracorporeal Membrane Oxygenation and Conventional Medical Therapy in Neonates with Persistent Pulmonary Hypertension of the New Born: A Prospective Randomised Study," *Pediatrics,* v. 84, (1989), pp. 957-963.

PAPINEAU, D., "The Virtues of Randomization," *The British Journal for the Philosophy of Science,* v. 45, n. 2, (1994), pp. 437-450.

PEARL, J., *Causality-Models, Reasoning and Inference,* Cambridge University Press, New York and Cambridge, 2000.

PENSTON, J., *Fiction and Fantasy in Medical Research: The Large-Scale Randomised Trial,* The London Press, London, 2003.

POCOCK, S. J., *Clinical Trials-A Practical Approach,* John Wiley, Chichester and New York, 1983.

POCOCK, S. J., "Statistical and Ethical Issues in Monitoring Clinical Trials," *Statistics in Medicine,* v. 12, (1993), pp. 1459-1469.

TUKEY, J. W., "Some Thoughts on Clinical Trials, especially Problems of Multiplicity," *Science,* v. 198, (1977), pp. 1958-1960.

WARE, J. H. and EPSTEIN, M. D., "Comments on 'Extracorporeal Circulation in Neonatal Respiratory Failure: A Prospective Randomized Study' by R. H. Bartlett et al.," *Pediatrics,* v. 76, (1985), pp. 849-851.

WORRALL, J., "What Evidence in Evidence-Based Medicine?," *Philosophy of Science,* v. 69, n. S3, (2002), pp. S316-330.

WORRALL, J., "Why There's no Cause to Randomize," *The British Journal for the Philosophy of Science,* forthcoming.

Prediction as Scientific Test of Economics

Wenceslao J. Gonzalez*

1. From Prediction as Scientific Test to its Role in the Duality Descriptive Economics-Normative Economics

Habitually, prediction is among those topics that receive an intense attention in economics, both from the viewpoint of the characterization of economic *undertaking* and from the angle of its incidence in determining whether economics is a *science* or not. De facto, prediction is a central issue in methodology of economics. This can be seen from internal criteria, such as the debate on prediction as scientific *test* of economics or the efforts to diminish forecast errors, and it can be perceived by external evaluations, such as the variety of specialized journals[1] or the number of Nobel Prize winners who have worked on the problem of prediction making methodological contributions. Among them are two recent economists laureate: Daniel Kahneman (in 2002)[2] and Clive Granger (in 2003).[3] They follow different lines of research: the former works on the relation with psychology, whereas the latter proposes models for the realm of statistical economics and econometrics.

At the same time, economic prediction is a central topic in the philosophic-methodological study of this discipline, because it has frequently been the axis of the debate on *reliability* of economic knowledge (and, hence, for its demarcation as scientific knowledge) in addition to being a key notion in the main *methodological controversies* in economics. Even though the origins of economics as science date back to Adam Smith, more than two hundred years ago,[4] the debate on its scientific status is

---

* The author is grateful to the Spanish Ministry of Science and Technology for supporting this work (research project HUM2004-06846/FISO).

1 Papers on economic predictions —from whatever angle— have a common presence in economic journals of general character. Besides these, there are a large number of journals devoted to issues on "prediction," "foresight," "planning," ...: *Quarterly Predictions of National Income and Expenditure*, *Journal of Forecasting*, *International Journal of Forecasting*, *Economics of Planning*, *Journal of Policy Modeling*, *Futures Research Quarterly*, *Technological Forecasting and Social Change*, *Journal of Time Series Analysis*…

2 Cf. Kahneman, D. and Tversky, A., "On the Psychology of Prediction," *Psychological Review,* v. 80, (1973), pp. 237-25l; Kahneman, D. and Tversky, A., "Prospect Theory: An Analysis of Decisions Under Risk," *Econometrica,* v. 47, (1979), pp. 313-327; Kahneman, D. and Snell, J., "Predicting Utility," in Hogarth, R. M. (ed), *Insights in Decision Making,* The University of Chicago Press, Chicago, 1990, pp. 295-310; and Kahneman, D., "Maps of Bounded Rationality: Psychology for Behavioral Economics," *The American Economic Review,* v. 93, n. 5, (2003), pp. 1449-1475.

3 Cf. Granger, C. W. J. and Newbold, P., *Forecasting Economic Time Series*, Academic Press, N. York, 1977; Granger, C. W. J., *Forecasting in Business and Economics*, 2nd ed., Academic Press, S. Diego, 1989 (1st ed., 1980); Granger, C. W. J. and Pesaran, M. H., "Economic and Statistical Measures of Forecast Accuracy," *Journal of Forecasting*, v. 19, (2000), pp. 537-560: Granger, C. W. J., "Evaluation of Forecasts," in Hendry, D. F. and Ericsson, N. R. (eds.), *Understanding Economic Forecasts*, The MIT Press, Cambridge (MA), 2002, pp. 93-103; and Granger, C. W. J. and Poon, S., "Forecasting Volatility in Financial Markets," *Journal of Economic Literature*, v. 41, (2003), pp. 478-539.

4 Smith, A., *An Inquiry into the Nature and Causes of The Wealth of Nations*, W. Strahan and T. Cadell, London, 1776. Edited by Edwin Cannan with preface by George J. Stigler, The University of Chicago Press, Chicago, 1976.

still going on.[5] Thus, one Nobel Prize winner —Sir John Hicks— holds that economics is not yet a "science," and it is not one because of the role played by prediction.[6]

When economists are dealing with scientific prediction in economics, they follow basically two kinds of methodological orientations: the broad position and the restricted perspective. The *broad position* seeks a nexus between methodology of economics and methodological options of general character (verificationism, falsificationism, methodology of scientific research programs...) or the comparison with the characteristic features of other disciplines in whatever scientific realm (formal, natural, social, or artificial). Meanwhile, the *restricted perspective* is focused on the specific case of economics itself. Thus, it does not try to compare its methodology with that used in natural sciences, social sciences,[7] or even sciences of the artificial.[8] The authors of economic theory, when they are making methodology, tend frequently to the broad position,[9] while the specialists in statistical economics and econometrics normally move towards the restricted perspective. The analysis in this paper will follow the first methodological orientation.

The problem of economic prediction also affects the way of understanding the relations between *two large branches* that traditionally conform economics: descriptive and normative.[10] Thus, the task of economic prediction in the first case —the descriptive or "positive" dominion— would have a different appearance from prediction in the second terrain (the normative or "political" facet), insofar as they have dissimilar missions and they possess a diverse practical repercussion for social life. This question can be analyzed in terms of the distinction between basic science and applied science,[11] where economic theory belongs to the former and statistical economics, as well as econometrics, fall into the latter.

2. The Problem of Prediction as Scientific Test of Economics

Terence Hutchison, as a historian of economic thought —and after half a century publishing papers—, reflects on the change of aims in economics, and he looks back on

[5] Cf. Rosenberg, A., *Economics-Mathematical Politics or Science of Diminishing Returns*?, The University of Chicago Press, Chicago, 1992; and Rosenberg, A., "La Teoría Económica como Filosofía Política," *Theoria*, v. 13, n. 32, (1998), pp. 279-299.

[6] Cf. Hicks, J., "A Discipline not a Science," in Hicks, J., *Classics and Moderns. Collected Essays on Economic Theory, v. III*, Harvard University Press, Cambridge, 1983, pp. 364-375; and Hicks, J., "Is Economics a Science?," in Baranzini, M. and Scazzieri, R. (eds.), *Foundations of Economics. Structures of Inquiry and Economic Theory*, B. Blackwell, Oxford, 1986, pp. 91-101.

[7] Cf. Gonzalez, W. J., "Marco teórico, trayectoria y situación actual de la Filosofía y Metodología de la Economía," *Argumentos de Razón Técnica*, v. 3, (2000), pp. 13-59.

[8] For Herbert Simon, economics is also a science of the artificial, cf. Simon, H. A., *The Sciences of the Artificial*, 3rd ed., The MIT Press, Cambridge, 1996, pp. 25-49; especially, p. 25.

[9] Cf. Sen, A., "Prediction and Economic Theory," in Mason, J., Mathias, P. and Westcott, J. H. (eds.), *Predictability in Science and Society*, The Royal Society and The British Academy, London, 1986, pp. 103-125.

[10] This distinction is frequently presented in terms of "positive economics" versus "political economics." As Terence W. Hutchison has pointed out, the explicit initial attempts to differentiate "positive" propositions of economic science from recommendations on economic policies to be followed, according to goals to be achieved, come from the first half of XIX century (i.e., John Stuart Mill and Nassau William Senior). Moreover, he recognizes that the distinction has not been always clear and that not all economists have considered this duality as an adequate focus to develop methodology, cf. Hutchison, T. W., *"Positive" Economics and Policy Objectives*, Allen and Unwin, London, 1964.

[11] On the distinction between "basic science" and "applied science," cf. Niiniluoto, I., "The Aim and Structure of Applied Research," *Erkenntnis*, v. 38, (1993), pp. 1-21.

what was already pointed out in XIX century around the centenary of the publication of *The Wealth of Nations* by Adam Smith. It was then that Robert Lowe, *Chancellor of the Exchequer* and who accepted a type of economy based on David Ricardo, stated the idea of prediction as *test* to determine the existence of science, and he assured that Smith satisfied —in the main— that condition.[12] Hutchison suggests implicitly a relevant distinction: prediction as *aim* of science, in general, and of economics, in particular; and prediction as *test* to determine the scientific character of economics. They are two different aspects that are not so explicitly distinguished in his presentation.

In his analysis, Hutchison is well aware of the relevance of the debate on prediction from an internal viewpoint. Thus, in one of the chapters of his book *Changing Aims in Economics*, entitled "to predict or not to predict? (That *is* the Question),"[13] he places us directly in front of the theme here discussed: whether we should accept that prediction —and, hence, predictive success— is the *test* to know whether economics is "science." Facing the problem of prediction as *requisite* for having science, there is a Hamletian doubt. The response to this issue has consequences for the general set of sciences and has a more intense repercussion in the case of economics, where the controversy is greater.

When the problem is whether economics is a "science," the issue of the role played by prediction immediately arises. This is not due to a mere influence of modern science, in general, and of Newtonian physics, in particular, which was looking for well grounded explanations and reliable predictions, but it has also more recent roots. In effect, Milton Friedman, in his famous paper *The Methodology of Positive Economics*, puts prediction as the *aim* or *goal* of economics.[14] It may be the echo of the philosophical and methodological approach developed by Hans Reichenbach, who considered prediction as the central aim of science,[15] but there is no indication of the direct influence of the leader of Berlin school on the mentor of economic monetarism.

Yet, the methodologist of "positive economics" and the author of *Experience and Prediction* subscribe an *instrumentalist* methodology:[16] both think that scientific process is conceived as a necessary *means* to an already-scheduled *end*, that is to make a prediction. Then, their main point of convergence lies in the subordination of scientific knowledge to the achievement of successes in predictions. In both authors there is a particular preference

---

[12] Cf. HUTCHISON, T. W., *Changing Aims in Economics*, B. Blackwell, Oxford, 1992, p. 72. Robert Lowe's text, originally published in *Political Economy Club* in 1876, says this in its page 7: "The test of science is prevision or prediction, and Adam Smith appears to me in the main to satisfy that condition."

[13] HUTCHISON, T. W., "To Predict or not to Predict? (That *is* the Question)," in HUTCHISON, T. W., *Changing Aims in Economics*, pp. 71-88 and notes in pp. 158-167.

[14] Cf. FRIEDMAN, M., "The Methodology of Positive Economics," in FRIEDMAN, M., *Essays in Positive Economics*, The University of Chicago Press, Chicago, 1953 (6th reprint, 1969), pp. 3-43. The first section analyzes "the relation between positive economics and normative economics." For him, "the conclusions of positive economics seem to be, and are, immediatedly relevant to important normative problems, to questions of what ought to be done and how any given goal can be attained," FRIEDMAN, M., "The Methodology of Positive Economics," p. 4.

[15] Cf. REICHENBACH, H., *Experience and Prediction*, The University of Chicago Press, Chicago, 1938.

[16] Cf. GONZALEZ, W. J., "Economic Prediction and Human Activity. An Analysis of Prediction in Economics from Action Theory," *Epistemologia*, v. 17, (1994), pp. 253-294; and GONZALEZ, W. J., "Reichenbach's Concept of Prediction," *International Studies in the Philosophy of Science*, v. 9, n. 1, (1995), pp. 35-56.

for empirical results rather than a predilection for the realm of theoretical assumptions.[17] Furthermore, they are in tune with fallibilists considerations: they accept the revision of scientific knowledge. That *epistemological fallibilism* adopted by Friedman and Reichenbach has different tones: in the former it has a "falsificationist" flavor —or, better, "refutationist"—, whereas the latter shows an indisputable empiricist component.[18]

2.1. Orientations in the Broad Position in Special Methodology of Economics: Four Divergent Options

The views differ among the specialists who deal with the methodology of economics in connection to the general methodology of science. At least, since 1980,[19] there is a clear interest in a methodological analysis of economics in those initially trained as economists.[20] But they follow two different lines: that characteristic of general methodological orientations of science, according to their diverse variants (Millian, verificationist, falsificationist, Kuhnian, Lakatosian, etc.), and that specific to the methodological controversies in economics (from Friedman's instrumentalism to McCloskey's rhetoric).[21]

De facto, the level of projection from the general methodological realm on the special sphere varies greatly. Moreover, rather than an effective integration of elements from general methodology of science (Millian, verificationist, falsificationist, Kuhnian, Lakatosian, etc.) in the concrete domain of economics, there is commonly a juxtaposition of contents: an appeal to the general realm at the beginning of some methodological approaches and a posterior development according to the characteristic mentality of the economists in question.[22]

Four economists are particularly relevant for their views on the problem of prediction in economic theory and its connection with issues of science, in general, and other

[17] There is a certain parallelism between Friedman and Reichenbach regarding the possibility of success in predictions that is not based on the realism of assumptions. This thesis is explicitly defended by Friedman and has been criticized strongly by other economists. (Cf. SIMON, H. A., "Testability and Approximation," in HAUSMAN, D. M. (ed.), *Philosophy of Economics*, Cambridge University Press, Cambridge, 1984 (1st ed.), pp. 245-248. Originally, it was published as SIMON, H. A., "Problems of Methodology-Discussion", *American Economic Review*, v. 53, (1963), pp. 229-231.) The analogous view would be the capacity of success that Reichenbach recognizes to the *clairvoyant* (cf. *Experience and Prediction*, p. 354), because a "clairvoyant" does not need any assumptions based on solid ground —theoretical and empirical.

[18] I am indebted to Wesley Salmon regarding the characterization of Hans Reichenbach within epistemological fallibilism.

[19] Cf. GONZALEZ, W. J., "Marco teórico, trayectoria y situación actual de la Filosofía y Metodología de la Economía," pp. 13-59; especially, pp. 15-18 and 37-56.

[20] On the last decade, cf. BACKHOUSE, R. E. (ed.), *New Directions in Economic Methodology*, Routledge, London, 1994; BOYLAN, TH. A. and O'GORMAN, P. F., *Beyond Rhetoric and Realism in Economics. Towards a Reformulation of Economic Methodology*, Routledge, London, 1995; BACKHOUSE, R. E., *Truth and Progress in Economic Knowledge*, E. Elgar, Cheltenham, 1997; and HANDS, D. WADE, *Reflection without Rules. Economic Methodology and Contemporary Science Theory*, Cambridge University Press, Cambridge, 2001.

[21] Cf. GONZALEZ, W. J., "From *Erklären-Verstehen* to *Prediction-Understanding*: The Methodological Framework in Economics," in SINTONEN, M., YLIKOSKI, P. and MILLER, K. (eds.), *Realism in Action*: *Essays in the Philosophy of Social Sciences*, Kluwer, Dordrecht, 2003, pp. 33-50.

[22] Cf. BLAUG, M., *The Methodology of Economics: Or How Economists Explain*, Cambridge University Press, Cambridge, 1980, pp. xi-xiii. This juxtaposition has been characterized as *methodological schizophrenia*, "in which methodological pronouncements and practice regularly contradict one another," HAUSMAN, D. M., *The Inexact and Separate Science of Economics*, Cambridge University Press, Cambridge, 1992, p. 7.

sciences, in particular. They are the Nobel Prize winners Milton Friedman (1976), John Hicks (1972), James Buchanan (1986), and Herbert A. Simon (1978). Each has a different philosophic-methodological conception of this problem. As I have shown elsewhere,[23] their outlooks can be characterized in the following terms: a) predictivist thesis; b) quasi-scientific option; c) dualist posture; and d) wary attitude.

2.2. Friedman's Predictivism

Undoubtedly, the *predictivist thesis* embraced by Friedman has received more attention than any other position in this regard. It has been very influential in economic theory for decades. It involves the primacy of prediction not only in methodological terms but also in axiological terms. Thus, prediction is —in his judgment— the main criteria of *scientific validity* and the *crucial goal* for science as a whole. In addition, it includes an epistemological stance where there is no discernible difference, regarding objectivity of knowledge, between natural realm and social domain.

Epistemologically, economics is —for Friedman— comparable to physics on the possibility of objective knowledge. This bottom convergence accompanies the methodological similitude about prediction. Thus, the task of positive economics would be "to provide a system of generalizations that can be used to make correct predictions about the consequences of any change in circumstances. Its performance is to be judged by the precision, scope, and conformity with experience of the predictions it yields. In short, positive economics is, or can be, an 'objective' science, in precisely the same sense as any of the physical sciences."[24]

Methodologically, "the only relevant test of the *validity* of a hypothesis is comparison of its predictions with experience. The hypothesis is rejected if its predictions are contradicted ('frequently' or more often than predictions from an alternative hypothesis); it is accepted if its predictions are not contradicted; great confidence is attached to it if it has survived many opportunities for contradiction. Factual evidence can never 'prove' a hypothesis; it can only fail to disprove it, which is what we generally mean when we say, somewhat inexactly, that the hypothesis has been 'confirmed' by experience."[25]

Axiologically, "the ultimate goal of a positive science is the development of a 'theory' or 'hypothesis' that yields valid and meaningful (…) predictions about phenomena not yet observed."[26] The method of an empirical science appears in Friedman as an *instrument* oriented toward such end. Thus, "viewed as a body of substantive hypotheses, theory is to be judged by its predictive power for the class of phenomena which is intended to 'explain' [sic]. Only factual evidence can show whether it is 'right' or 'wrong' or, better, tentatively 'accepted' as valid or 'rejected'."[27]

---

[23] Cf. GONZALEZ, W. J., "On the Theoretical Basis of Prediction in Economics," *Journal of Social Philosophy*, v. 27, n. 3, (1996), pp. 201-228; and GONZALEZ, W. J., "Prediction and Prescription in Economics: A Philosophical and Methodological Approach," *Theoria*, v. 13, n. 32, (1998), pp. 321-345.

[24] FRIEDMAN, M., "The Methodology of Positive Economics," in FRIEDMAN, M., *Essays in Positive Economics*, p. 4.

[25] FRIEDMAN, M., "The Methodology of Positive Economics," pp. 8-9.

[26] "The Methodology of Positive Economics," p. 7

[27] FRIEDMAN, M., *Ibidem*, p. 8.

2.3. Hicks's Quasi-scientific Option

Quite different is the *quasi-scientific option* defended by Hicks. Semantically, he dismisses that economics can be called "science," because he thinks that it is only a "discipline" which is on the edge of science and, therefore, it did not get so far as the scientific status.[28] In effect, he considers that this subject is a discipline that is "on the edge of science, because it can make use of scientific, or quasi-scientific, methods; but it is no more than on the edge, because the experiences that it analyses have so much that is non-repetitive about them."[29]

From the methodological point of view, Hicks maintains that is not possible to obtain a confluence of economics and physics, because he holds that they belong to two terrains of reality that are neatly different. Thus, the kind of knowledge is diverse in those areas: "the facts which we study [in economics] are not permanent, or repeatable, like the facts of natural sciences; they change incessantly, and change without repetition."[30] In his judgment, physics works on repeatable phenomena and, consequently, deals with predictable facts; whereas economics, insofar as is focused on ever-changeable events, has the problem of predictability, due to the difficulty in establishing a connection between the initial stage and the posterior situation.

Hicks, through the "quasi-scientific" option, does not accept a methodological convergence between natural science —especially, physics— and social sciences (mainly, economics), because he considers them as clearly split by the problem of prediction. In this regard, he draws frontiers regarding the kinds of predictions: in physics there are *unconditional* prediction (it establishes that something will happen in a certain way) and *strong conditional* prediction (where something will occur if previously a condition is fulfilled);[31] whereas in economics the dominant is *weak conditional* prediction, that "it says no more than the event will follow, if there is no disturbance."[32] In his judgment, economic prediction are commonly weak predictions, they take the form of 'this is what will happen, if something does not come up to prevent it.' This weakness is subject to a *ceteris paribus* (other things being equal) clause.[33]

Against the predictivist thesis, Hicks rejects that the dominant cognitive value in economics may be the predictive success. Hence, from an axiological viewpoint, he does not accept that the ultimate goal of this discipline could be to predict, because economics receives stronger influence from its past —its history— than from his future. He considers that "economics is more like art or philosophy than science, in the use that it can make of

---

[28] Cf. Hicks, J., "A Discipline not a Science," in Hicks, J., *Classics and Moderns. Collected Essays on Economic Theory*, *v. III*, pp. 364-375.

[29] Hicks, J., "Is Economics a Science?," in Baranzini, M. and Scazzieri, R. (eds.), *Foundations of Economics. Structures of Inquiry and Economic Theory*, p. 100.

[30] Hicks, J., "'Revolutions' in Economics," in Hicks, J., *Classics and Moderns. Collected Essays on Economic Theory, v. III*, p. 4.

[31] In this conception, "unconditional" does not mean without "initial conditions:" it expresses something that runs on its own; meanwhile "conditional" expresses the need of the existence of some phenomena that could or not happen.

[32] Hicks, J., "Is Economics a Science?," p. 94.

[33] Cf. "Is Economics a Science?," p. 97.

its own history."[34] Economic predictions are possible, they may even be of different kinds, but normally they are weak.

2.4. Buchanan's Dualism

Through the *dualist posture*, Buchanan offers a methodological orientation to economics that differs from the predictivist thesis and the quasi-scientific option. In his main book in this regard —*Economics: Between Predictive Science and Moral Philosophy*—[35], he maintains that there are two different methodological realms in economics: the objective sphere and the subjective area. The first territory includes predictions that can be scientific, whereas the second domain involves non-predictive knowledge. Furthermore, he points out an additional factor of historical character: economics starts with Adam Smith as a discipline related to moral philosophy.

Explicitly, Buchanan accepts a "subjective economics,"[36] whose content would be "defined precisely within the boundaries between the positive, predictive science of the orthodox model on the one hand and the speculative thinking of moral philosophy on the other."[37] This subjective economics is not oriented towards prediction. Thus, the role of prediction would be confined to the other sector of economics, that part whose knowledge can be qualified as "scientific." Therefore, Buchanan presents a *tertium quid* in comparison with the previous approaches: his view is in tune with the quasi-scientific option insofar as he accepts a territory of economics that is not scientific, whereas he is closer to the predictivist thesis when he assumes that there is a part of economics —"positive"— that can make scientific predictions.

A semantic as well as an epistemological frontier would distinguish these two camps of economics. Thus, the discourse of (non-predictive) subjective economics "can offer insights into the dynamics through which a society of persons who remain free to choose in a genuine sense develops and prospers."[38] Meanwhile the language of (predictive) positive economics will refer to those events which are more obvious in human interaction. Moreover, subjective economics "occupies an explanatory realm that is mutually exclusive with that properly occupied by positive economics."[39] The domain of knowledge of "subjective economics" would be located between the terrain of empirical science, where there are scientific tests —which include predictions—, and moral philosophy, which is certainly not conceived of as a predictive discipline (even though it may be, partly, prescriptive and there exists a relation between predictions and prescriptions).

---

34 HICKS, J., "'Revolutions' in Economics," p. 4.

35 BUCHANAN, J. M., *Economics: Between Predictive Science and Moral Philosophy*, Texas A & M University Press, College Station, 1987.

36 "Economists often complain about the observed fact 'everyone is his own economist,' in an expression of the view that scientific counsel fails to command the deference it seems to warrant," BUCHANAN, J. M., "Economics as a Public Science," in MEDEMA, S. G. and SAMUELS, W. J. (eds.), *Foundations of Research in Economics: How do Economists do Research*, E. Elgar, Brookfield, VT, 1996, p. 34.

37 BUCHANAN, J. M., "The Domain of Subjective Economics: Between Predictive Science and Moral Philosophy," in BUCHANAN, J. M., *Economics: Between Predictive Science and Moral Philosophy*, p. 68.

38 BUCHANAN, J. M., "The Domain of Subjective Economics: Between Predictive Science and Moral Philosophy," p. 70

39 "The Domain of Subjective Economics: Between Predictive Science and Moral Philosophy," p. 70.

Buchanan subscribes a *methodological dualism* within economics itself, because he considers that this distinction (objective/predictive and subjective/non-predictive) is necessary in order to avoid confusion when the history of economic doctrines is made. Positive economics requires a specific methodology: what can (conceptually) be predicted, it may be —in his judgment— explained with an objective or scientific theory; whereas this does not happen when something cannot be predicted, because then it is only possible explain (or understand) the event through a subjective theory. Furthermore, he points out that, in the application of scientific knowledge —political economy—, categorical differences with other sciences can be seen[40] (especially, in the cases of physics, chemistry and biology).

2.5. Simon's Wary Attitude

Following a critical approach to mainstream economics, Simon offers a *wary attitude* towards the role of prediction. His views are different from the previous ones (the predictivist thesis, the quasi-scientific option, and the dualist posture). He suggests putting aside the issue of prediction as test of the scientific character of economics: "we should be wary of using prediction as a test of science, and especially of whether economics is a science, for an understanding of mechanisms does not guarantee predictability."[41]

However, Simon includes prediction as a central topic in the *definition* of what economics is: "*economics* can be defined as the science that describes and predicts the behavior of the various kinds of economic man."[42] His stance is pragmatic and connects with the links between prediction and prescription: we need predictions to handle future economic events (and then resolve concrete problems); and, at the same time, economists "can even provide theoretical reasons why it should be impossible to predict the course of the business cycle or the stock market."[43]

Instead of emphasizing precision in the economic results (i.e., predictive success), like Friedman's instrumentalism, Simon highlights the *understanding* of the mechanisms: to grasp the *procedures*, rather than to achieve fine results. For him, the focus is not in guaranteeing predictability: to predict is neither the only aim of economics nor the central goal of this science. Moreover, he seems to change the priorities of the predictivist thesis: economics can be evaluated through the assessment on the correction of the assumption instead of being tested by means of the empirical adequacy of its predictions. This viewpoint is a consequence of his insistence on the need to start searching for *facts*.[44]

---

[40] "It is only in the application of its scientific content that economics comes to be categorically distinct from its scientific neighbors," Buchanan, J. M., "Economics as a Public Science," in Medema, S. G. and Samuels, W. J. (eds.), *Foundations of Research in Economics: How do Economists do Research*, p. 32.

[41] Simon, H. A., "The State of Economic Science," in Sichel, W. (ed.), *The State of Economic Science. Views of Six Noble Laureates*, W. E. Upjohn Institute for Employment Research, Kalamazoo, MI, 1989, p. 100.

[42] Simon, H. A., "Economics and Psychology," in Koch, S. (ed.), *Psychology: A Study of a Science*, vol. 6, McGraw Hill, N. York, 1963. Reprinted in Simon, H. A., *Models of Bounded Rationality, vol. 2: Behavioral Economics and Business Organization*, The MIT Press, Cambridge, 1982, p. 320.

[43] Simon, H. A., "The State of Economic Science," p. 99.

[44] "The faith in *a priori* theory, uncontaminated by empirical observations, has been weakened —even among 'rational expectationists.' More and more economists are beginning to look for the facts they need in actual observation of business decision making and in laboratory experiments on economic markets and organizations," Simon, H. A., "Introductory Comment," in Egidi, M. and Marris, R. (eds.), *Economics, Bounded Rationality and the Cognitive Revolution*, E. Elgar, Aldershot, 1992, p. 7.

Semantically, Simon accepts "prediction" in the sense of anticipation of future events based on present knowledge, which has as reference the complex economic world that, in principle, is ruled by uncertainty. But, in his approach, *prediction* is not linked to a necessary component of the future: it also can be a testable implication without the temporal factor. However, he distinguishes clearly the cognitive realm of "prediction" and the practical activity of "planning."[45] From a logical point of view, in the relations between "explanation" and "prediction,"[46] Simon seems to follow an asymmetrical pattern rather than a symmetrical path, insofar as he criticizes the possibility of a theory of decision-making understood as a nomological theory with strict predictions (uncertainty is incompatible with mere deductive inferences towards future). Furthermore, he establishes differences between models of natural sciences and social sciences.[47]

Simon insists on the role of experience as epistemological grounding of prediction in economics: he rejects whatever vision of an *a priori* knowledge and defends the need for objectivity (which leads him towards the respect for reality: uncertainty, bounded rationality, etc.). Epistemologically, his views sometimes tend to be close to Logical empiricism,[48] whereas on other occasions he is very critical of that conception. Methodologically, prediction has a well defined structure in his approach: i) the insufficiency of pure deductions leaves the open door to *induction*; ii) the need for realism of assumptions acts in favor of a *non-instrumentalist* perspective; and iii) the importance of interdisciplinary work (economics connected to psychology as well as computer science —artificial intelligence— and other disciplines) gives a broad margin for economic prediction within the social sciences.[49]

Generally, Simon maintains a critical attitude towards mainstream economics (the neoclassical conception) in key points. On the one hand, regarding the role of prediction in economic theory, he rejects that prediction should be its main *aim* or central goal. Furthermore, he is uncomfortable with the *assumptions* of economic models, such as those proposed by Friedman,[50] which require substantive rationality and emphasize successful prediction as decisive. On the other hand, concerning the role of prediction as the *test* of

[45] "Our practical concern in planning for the future is what we must do *now* to bring that future about. We use our future goals to detect what may be irreversible present actions that we must avoid and to disclose gaps in our knowledge (…) that must be closed soon so that choices may be made later," SIMON, H. A., "Prediction and Prescription in Systems Modeling," *Operations Research*, v. 38, (1990), p. 11.

[46] The topic of symmetry and asymmetry of explanation and prediction has a long tradition in philosophy of science. Even authors originally formed in Logical empiricism have defended later on the asymmetry thesis, cf. SALMON, W., "On the Alleged Temporal Anisotropy of Explanation," in EARMAN, J., JANIS, A. MASSEY, G. and RESCHER, N. (eds.), *Philosophical Problems of the Internal and External Worlds*, University of Pittsburgh Press, Pittsburgh, 1993, pp. 229-248.

[47] Cf. SIMON, H. A., "Rational Decision Making in Business Organizations," *American Economic Review*, v. 69, n. 4, (1979), p. 510.

[48] "Assumptions to be supported by publicly repeatable observations that are obtained and analyzed objectively," SIMON, H. A., "Rationality in Psychology and Economics," in HOGARTH, R. M. and REDER, M. W. (eds.), *Rational Choice. The Contrast between Economics and Psychology*, The University of Chicago Press, Chicago, 1987, p. 28.

[49] Cf. GONZALEZ, W. J., "Rationality in Economics and Scientific Predictions: A Critical Reconstruction of Bounded Rationality and its Role in Economic Predictions," *Poznan Studies in the Philosophy of Science*, v. 61, (1997), pp. 205-232; especially, pp. 219-221.

[50] Cf. FRIEDMAN, M., "The Methodology of Positive Economics," pp. 3-43; especially, pp. 16-30.

economics as a science, his choice is a wary attitude towards prediction as the touchstone to decide on the scientific character of economics.[51] His interest is oriented to understanding the mechanism —the processes— that explain past phenomena and present events rather than the predictability of economic behavior.

Nevertheless, according to Daniel Hausman,[52] it is possible to distinguish three different views on prediction: (a) as testable implications regarding the future, (b) as testable implications whose truth is not already known, and (c) as testable implications. According to this analysis, Simon does not deny the importance of prediction in sense (c) as the central test of science, but he denies that prediction in sense (a) is the crucial test of economics as a science. To put it differently: economics needs testable implications obtained through economic processes of the past and the present, whereas it should not be tested taking as cornerstone the pure testable implications regarding the future. In other words: the priority is *to understand* economic processes rather than to predict them. Thus, even though a prediction may be correct or successful, the important factor is —for him— to grasp the mechanisms that have led to that result or outcome.

3. Analysis of Philosophic-Methodological Options to Economics Regarding Prediction and its Role as Scientific Test

Until now, the focus of attention has been on the approaches of relevant economists that, when reflecting on their discipline, try to clarify the role of prediction within the framework of the scientific status of economics. Their viewpoints on methodology can be analyzed taking into account the contributions that philosophers have made within the general methodological realm. These considerations can be used as a setting to evaluate the approaches developed by the economists about economic prediction and its role as scientific test. This link involves an interconnection of the general methodology of science and the methodology of economics of broad position.

Historically, general methodology of science has had variations in the recognition of the role of prediction as scientific test. On the one hand, the relevance of prediction as *scientific value* —a relevant aim— and a central factor for *evaluation* of scientific theories is commonly accepted. But on the other, there is an insistence on the *accommodation* regarding known facts, which could be at least as important as successful scientific *prediction*. This backdrop contributes to seeing more clearly the problem that arises when the general vision is projected onto the particular case of economics.

3.1. Methodological Orientations of General Character

The discrepancies in the characterization of prediction can be seen through the diverse methodological orientations. From Hans Reichenbach to Philip Kitcher,[53] it can be pointed out that, except Logical positivists at the beginning of the movement (such as Rudolf

[51] Cf. Simon, H. A., "The State of Economic Science," in Sichel, W. (ed.), *The State of Economic Science. Views of Six Nobel Laureates*, p. 100.

[52] Cf. Hausman, D., *Personal Communication*, 21 January 1996.

[53] Cf. Kitcher, Ph., *Science, Truth and Democracy*, Oxford University Press, N. York, 2001, pp. 5, 8, 16-18, 20-24, 28, 30, 33-36, 40, 80, 82, 98, 103, 154, 177-180, 182-183, 185, 192, and 194-195.

Carnap in the period of the Vienna Circle),[54] most philosopher of science —those more relevant— have reflected, more or less intensely, on the role or prediction in science.[55] The reason is clear: science also works on statements about the future which should be testable. Ordinarily, they have no difficulties in accepting that prediction may be an *aim* of scientific undertaking —a pertinent cognitive value—, and habitually they admit that prediction has an appropriate role as scientific *test*. Nevertheless, the emphasis of each of the main philosophic-methodological conceptions varies.

Reichenbach represents a clear predictivism in the general methodology of science: he conceives science as directed towards predictive success. Karl Popper highlights the role of prediction for science in general,[56] but he criticizes scientific social prediction of long run and large extent. However, even in the terrain of the social sciences he does not deny the role of prediction as scientific *test*. De facto, he explicitly accepts it for testing economic theories.[57] Imre Lakatos stresses prediction as a feature of a progressive research program, and he has made contributions in the context of prediction of *novel facts*.[58]

Other authors engaged in the "historical turn," such as Thomas Kuhn or, later on, Larry Laudan —in his two periods—, include prediction as a relevant factor of science. Kuhn, in his *Postscript-1969* to *The Structure of Scientific Revolutions*, maintains that deeply held scientific values are those concerning predictions, because they should be accurate.[59] Laudan, especially in *Progress and its Problems*,[60] incorporates prediction as one element more in the problem solving; and, in the posterior phase of "normative naturalism," through his theory of comparative evaluation of scientific theories based in the evidence, he analyzes historical examples of predictions.[61] This allows that important methodological conceptions directly linked to the history of science also accept a vinculum between "scientific progress" and "prediction."

---

[54] Carnap's interest in this topic is clearly posterior to *Experience and Prediction*, the important book written by Reichenbach, cf. Carnap, R., "Theory and Prediction in Science," *Science*, v. 104, n. 2710 (6 December 1946), pp. 520-521.

[55] The most systematic general philosophic-methodological analysis can be seen in Rescher, N., *Predicting the Future*, State University Press New York, N. York, 1998.

[56] Cf. Popper, K. R., *The Logic of Scientific Discovery*, Hutchinson, London, 1959, p. 65, note 1. On his views on scientific prediction, cf. Gonzalez, W. J., "The Many Faces of Popper's Methodological Approach to Prediction," in Catton, Ph. and Macdonald, G. (eds.), *Karl Popper: Critical Appraisals*, Routledge, London, 2004, pp. 78-98.

[57] The argument against historicism "it is perfectly compatible with the possibility of testing social theories —for example, economic theories— by way of predicting that certain developments will take place under certain conditions," Popper, K., *The Poverty of Historicism*, Routledge and K. Paul, London, 1957, p. vii.

[58] Cf. Gonzalez, W. J., "Lakatos's Approach on Prediction and Novel Facts," *Theoria*, v. 16, n. 3, (2001), pp. 499-518; especially, pp. 505-508.

[59] "Probably the most deeply held values concern predictions: they should be accurate; quantitative predictions are preferable to qualitative ones; whatever the margin of permissible error, it should be consistently satisfied in a given field; and so on," Kuhn, Th. S., "Postscript-1969," in Kuhn, Th. S., *The Structure of Scientific Revolutions*, The University of Chicago Press, Chicago, 2nd ed., 1970, p. 185.

[60] Laudan, L., *Progress and its Problems*, University of California Press, Berkeley, 1977.

[61] Cf. Laudan, L., "Una Teoría de la evaluación comparativa de teorías científicas," in Gonzalez, W. J. (ed.), *El Pensamiento de L. Laudan. Relaciones entre Historia de la Ciencia y Filosofía de la Ciencia*, Publicaciones de la Universidad de A Coruña, A Coruña, 1998, pp. 155-169; especially, pp. 166-167.

As to the task of prediction as test of science, it seems patent that the role is recognized in those conceptions focused on the *content* itself of science and, hence, adopts an internal perspective on science. Moreover, the role of prediction as test is emphasized by those viewpoints that conceive scientific content as objective, as can be seen in Popper and, to a greater degree, in Lakatos. Meanwhile, the relevance of the predictive function seems diminished when the psycho-sociological dimension of science (or even purely sociological view) is highlighted. Predicting then is not a genuine test, but rather an activity of the scientific community, which should be evaluated according to the changing criteria assumed periodically by that community. Thus, moving towards a full-fledged external perspective of science, prediction can be included in the problems of the underdetermination of theories by experience or, what is equivalent, can go to a network of methodological relativism, where *de facto* there are no proper tests for science.

But the debate on prediction as test of scientific character of economics is raised in different terms. Habitually, it reflects a problem of limits as "barriers" (*Schranken*), which —according to Immanuel Kant— distinguish the safe way of science and the non-scientific orbit, instead of being discussed in terms of "confines" (*Grenzen*) of scientific activity (i.e., the ceiling or terminal limit that can be reached by science).[62] Thus, the issue of prediction as a requisite for science can arise within the "barriers" —the specific realm of scientific activity as different from other domains— and only afterwards may it be studied in the context of "confines."

Our science, insofar as it is *our science* (i.e., a human activity developed in a contemporary society), is open to the future. Moreover, due to the "principle of proliferation of questions" (where each response given raises a new question which needs to be answered), it is problematic to foretell what contents science will have after several decades; there exists a congenital difficulty in predicting future science with some reliability, because of the emergence of new issues in domains already explored and the discovery of previously unknown zones which lead to new questions.[63]

Accordingly, if it is assumed that science is lacking clear "terminal limits" and, at the same time, that it is always open to a progressive revision, then it is possible to accept that, in principle, scientific prediction can in the future improve its level of reliability as soon as there exists an enlargement of our knowledge of the variables within a specific domain. Thus, it is the case that, usually, the basic questions on prediction as *scientific test* arise from the philosophic-methodological analysis of prediction within the "barriers" of science. This topic is commonly discussed under the heading of the problem of "demarcation," which requires distinguishing between the complex phenomenon of science and what it is not science.

Within this discussion should be considered the constitutive elements of a science (language, structure, knowledge, method, dynamic activity in a social setting, aims, and

---

[62] On this distinction based on Kant, cf. Radnitzky, G., "The Boundaries of Science and Technology," in: *The Search for Absolute Values in a Changing World*, *Proceedings of the VIth International Conference on the Unity of Sciences*, International Cultural Foundation Press, New York, 1978, vol. II, pp. 1007-1036.

[63] This might lead to the unpredictability of future predictions, which could be the absence of confines —terminal limits— in science. Rescher sees the problem from the point of view of the impossibility —in his judgment— of present science saying how the future science will be, cf. Rescher, N., *The Limits of Science*, revised edition, University of Pittsburgh Press, Pittsburgh, 1999, Ch. 7, pp. 95-110.

values) insofar as they are different from the features of other kinds of disciplines or human activities.[64] Among these traits can be highlighted the fact of its being a human activity that follows an ordered procedure to increase knowledge. The possession of a *method* is what makes it possible for the knowledge to progress. This methodological side accompanies epistemology and the other facets. Especially relevant among them are semantics of science and logic of science as well as axiology of research. Scientific prediction, in general, and economic prediction, in particular, cannot ignore the contributions made at those philosophical levels.

When the issue of prediction is raised in terms of *requisite* of a science (i.e., as aim and test to determine if something is scientific or not), the initial question —to predict or not to predict— should be diversified according to the constitutive elements of a science. More relevant here are the methodological component and the epistemological grounding. If predicting is consubstantial to the fact itself of science, then prediction should appear in each of the ingredients of that complex reality; especially, in the methodological process, which serves to warrant the reliability and advancement of scientific knowledge.

However, it may be that prediction is not an indispensable element of a science as such (i.e., a necessary aim or goal) and, nonetheless, it can be used as test to guarantee that something is "scientific" instead of "non-scientific (or, at least, more scientific than other conceptions that do not include predictions or do not have success with the predictions that they made). This second approach would conceive prediction as factor of *demarcation* understood in a *weak* sense, while the first proposal, insofar as it sees prediction as consubstantial to scientific undertaking, would configure prediction as an aspect for demarcation in a strong sense. Hence, an analysis of every constitutive element of science (language, structure, knowledge, method, dynamic activity in a social setting, aims, and values) according to the role of prediction is needed.

After the philosophic-methodological analysis of those components, can prediction be *a test* of science? Without doubt, as general methodology of science has pointed out; but it cannot be, in principle, *the only* test or the *crucial evaluator* of scientific statements. Prediction is a requisite of any science (and, therefore, of economics insofar as it is a science), but it does not commonly represent a constitutive factor of science in itself: not every scientific theory should be predictive. Thus, it can demarcate in a weak way: prediction is a test that guarantees the scientific character of the method used. However, in order to have science, it is not indispensable to count on the existence of predictions.

Moreover, if the proposal were that prediction has a strong role to demarcate, then the consequences would be highly problematic: they might involve the movement from the present "barriers" between science and not science to positions that would exclude conceptions habitually considered as scientific. This is the case of the theory of evolution by variation and natural selection, proposed by Charles Darwin, which was not originally conceived with a predictive perspective (e.g., to describe new future species based on the available ones). It is normally accepted that this theory does not require necessarily the existence of a predictive component as a specific element in order to be scientific. Thus, even though this biological theory does not make explicit predictions about the future of

[64] Cf. GONZALEZ, W. J., "The Philosophical Approach to Science, Technology and Society," in GONZALEZ, W. J. (ed.), *Science, Technology and Society: A Philosophical Perspective*, Netbiblo, A Coruña, 2005, pp. 3-49.

actual species or does not foretell the coming into existence of creatures of some novel species, Darwin still figures among the best known scientists.[65]

3.2. Prediction as a Test in Economics and Philosophic-Methodological Options

From the philosophic-methodological criteria already indicated, the question raised by Hutchison —to predict or not to predict?— is answered in favor of the first possibility, although not exclusively: prediction is a *sufficient condition* for a science, not a necessary condition. Furthermore, that use of prediction —as a test— is not the only one employed in science, because it has also a dimension directly related to *prescription*,[66] especially in economics, where besides a "descriptive" (or "positive") branch there is another realm: "normative" (or "applied"). The different role of prediction is clear: in the first case it belongs to the attempt to *describe* the economic activity in the real setting (the agents in the transactions of goods and services, supply and demand, …), looking at the future, whereas in the second case prediction is put to the service of prescriptions, which seek the *desirable* courses of action to achieve some specific ends.[67]

Consequently, prediction appears in economics with two methodological tasks: a) as *a tool for testing* theories, in general, and hypotheses, in particular; and b) as *an instrument for public policy* (i.e., the direction of human actions in the social world, which includes political economy). In effect, if *descriptive economics* seeks to reflect the actual economic undertaking (the performance of agents in the world of commodities and services, supply and demand) and *normative economics* offers patterns to prescribe better economic choices and the desirable courses of action (what affects how an ideally rational agent may behave), then it would be a first modulation of the role of economic prediction. Predicting the "real" path of economic events and predicting the "ideal" route of economic endeavors are on two different levels, insofar as normative economics looks, in principle, to optimization (or to maximization, according to neoclassic economics) while descriptive economics does not seek that goal directly.

Predicting is a habitual *aim* of science, understood in the contemporary sense, and it is also a *test* to determine whether something is scientific or not. The different components of science (language, structure, knowledge, ...) have a relation to future, both in the territory of natural sciences and in the dominion of social sciences. When prediction is indicated here as a non-necessary condition for having "science," the analysis involves a rejection of Hick's extreme thesis (that of economics being only a discipline), because the consideration of the *general* components of science does not allow us to state that economics is merely a "discipline." As scientific test, the capacity of prediction varies, due to its relation with the

65 In this regard, Toulmin maintains that many of the most powerful theories did not state "verifiable forecasts," like the Darwinian theory of evolution by means of variation and natural selecction. On this issue, cf. TOULMIN, S., *Foresight and Understanding. An Enquiry into the Aims of Science*, Indiana University Press, Bloomington, 1961, pp. 24-27; especially, p. 24.

66 Within the economic field, the uses of prediction as a scientific test and as an element conected with prescription are emphasized by Herbert A. Simon elsewhere, especially in SIMON, H. A., "Prediction and Prescription in Systems Modeling," pp. 7-14. (Reprinted in SIMON, H. A., *Models of Bounded Rationality*. Vol. 3: *Empirically Grounded Economic Reason*, pp. 115-128.)

67 On the relations between "prediction" and "prescription," cf. GONZALEZ, W. J., "Prediction and Prescription in Economics: A Philosophical and Methodological Approach," pp. 329-339.

actual content related to it. Thus, prediction has a different degree of difficulty according to the *object* studied (number of variables considered), the *level of detail* of its content (the accuracy and precision) that is aimed and the boundaries of *time* (short, middle or long run) involved.

3.2.1. Preeminence of Prediction: Friedman's Predictivist Thesis

Considering these elements, Friedman's predictivist thesis, where the important factor is the success of prediction when tested by experience, goes too far. He has received a large amount of criticism, of which two points are relevant here. a) This thesis is seen as too radical: neither scientific theories, in general, nor economic theories, in particular, can commonly be reduced to the central goal of making predictions on non-observed phenomena. b) The thesis is built up on the basis of the explicit acceptance of a lack of realism in the assumptions,[68] because he insists on the relevance of the success itself in the prediction, where the validity of the process can be accepted even on non-realistic bases (which opens the door to false assumptions). Both aspects —the subordination of scientific character to the fulfillment of predictive goal and the acceptance as adequate of the lack of realism of assumptions— seem avoidable in a reasonable view on prediction.

Besides prediction, an important part of the methodology of economics seeks to accommodate to real facts, in order to give an "explanation" or to look for its "understanding." To circumscribe economics to the predictive task would be an unnecessary limitation, because prediction does not cover the whole field of this science. In effect, as Amartya Sen has pointed out, "prediction is not the only exercise with which economics is concerned. Prescription has always been one of the major activities in economics, and it is natural that this should have been the case. Even the origin of the subject of political economy, of which economics is the modern version, was clearly related to the need for advice on what is to be done in economic matters. Any prescriptive activity must, of course, go well beyond pure prediction, because no prescription can be made without evaluation and an assessment of the good and the bad."[69]

Along with this objection, it is possible to add another regarding the *criterion* itself chosen by Friedman (that of *success* in predicting, comparing what is predicted —theory or hypothesis— with the results of evidence), because a very frequent criticism of economic predictions is their unreliability. Thus, what the economists have predicted should be compared with what happens in fact (i.e., what is achieved from evidence). And, in consonance with the results obtained by this instrument, it is possible to argue as "Professor R. Clower has written: 'If successful prediction were the sole criterion of a science, economics should long have ceased to exist as a serious intellectual pursuit'."[70] Consequently, the methodological instrumentalism of the predictivist thesis is de facto very problematic.

---

68 "Truly important and significant hypotheses will be found to have 'assumptions' that are widly inaccurate descriptive representations of reality, and, in general, the more significant the theory, the more unrealistic the assumptions," FRIEDMAN, M., "The Methodology of Positive Economics", p. 14.

69 SEN, A., "Prediction and Economic Theory," in MASON, J., MATHIAS, P. and WESTCOTT, J. H. (eds.), *Predictability in Science and Society*, p. 3.

70 HUTCHISON, T. W., "On Prediction and Economic Knowledge," in HUTCHISON, T. W., *Knowledge and Ignorance in Economics*, Blackwell, Oxford, 1977, p. 12.

Another aspect questioned in Friedman's preditivism is that economic theory can produce sufficiently precise predictions even though there is no realism in the economic assumptions.[71] Among the critics is Simon, who calls for *realism of assumptions* in the theories on economic *actors* ("microeconomics") as well as in theories on economic *markets* ("macroeconomics"),[72] and in both cases he includes uncertainty about the future. The bases for prediction should be solid. The realism of assumptions is very important when the goal is to describe economic activity and to contribute subsequently to the prescriptions to resolve the concrete problems of economics.

It is beyond doubt that economics needs to make predictions, especially when it is required to offer prescriptions. But not every economic reality is related to producing predictions. Thus, besides making predictions, economics deals with other aims, such as the *description* of economic activity (mainly in the area of welfare economics) or the *evaluation* of results. Making predictions is then a relevant task that should be accompanied by other aspects, such as the description of human activity in the economic field or the evaluation of the result that it is convenient to obtain. Furthermore, testing a theory through its predictions is, frequently, a difficult task:[73] part of the pertinent information might not be available; the control of the observational process is sometimes a problematic issue; and it can raise a wide debate dealing with the available data (as is usual in issues related to unemployment or economic crisis).

3.2.2. Weakness of Prediction: Hicks's Option

To some extent, Hicks constitutes the extreme opposite to Friedman, insofar as he does not see economics as a "science" and puts economic predictions at a low level of reliability. Without going into details on his methodology of economics, a task already performed by Mark Blaug,[74] it is possible to analyze how Hicks conceives economic predictions. He maintains that there are two main kinds: on the one hand, predictions of what will happen; and, on the other hand, predictions of what might happen.[75] Predictions of physics and economics can be examples of these kinds. In his judgment, an economic prediction only can state what value a variable will have at a future date: it is the maximum aspiration. Expectations of future are formed as based for ordinary economic action.[76]

According to Hicks, some sciences are able to make some sorts of unconditional predictions, whereas the normal type of scientific prediction is the conditional one.[77]

---

[71] Cf. Friedman, M., "The Methodology of Positive Economics," p. 15.

[72] Cf. Simon, H., "Testability and Approximation," in Hausman, D. (ed.), *Philosophy of Economics*, 1st ed., pp. 245-248.

[73] Cf. Caldwell, B., *Beyond Positivism: Economic Methodology in the Twentieth Century*, Allen and Unwin, London, 1982; revised edition, Routledge, London, 1994, p. 174.

[74] After studying Hick's ideas on economics as a non-scientific "discipline," realism as a desideratum of economic models, the distinction between positive economics and normative economics, and the problem of causality, Blaug concludes: "it is impossible to extract any coherent methodology of economics from the writings of Hicks", Blaug, M., "John Hicks and the Methodology of Economics," in Marchi, N. (ed), *The Popperian Legacy in Economics*, Cambridge University Press, Cambridge, 1988, p. 194.

[75] Cf. Hicks, J., "A Discipline not a Science," p. 369.

[76] Cf. Hicks, J., *Ibidem*.

[77] Cf. Hicks, J., "Is Economics a Science?" pp. 93-94.

Unconditional predictions —that something will happen in a certain way— can be made in astronomy, because the astronomer can tell us, with remarkable precision, just when there will be an eclipse of the sun. Moreover, the phenomena studied by the astronomer are beyond the range of human influences; and this scientist can circumscribe his or her description of the phenomena, and can feel sure that there are no conditions that were not taken into account.

Within the field of conditional prediction, Hicks establishes a distinction between *strong* and *weak* predictions. "If it is strong, it states that, given the stated conditions, the event will follow; if it is weak, it says no more than that the event will follow, if there is no disturbance."[78] Thus, *conditional weak prediction* requires only that some of the conditions for the event to occur have been identified, but there are others that we cannot specify or cannot enumerate yet. He claims that economic predictions are generally weak predictions. They are based on weak propositions and take the form 'this is what will happen, if something does not come up to prevent it.' The weakness is in it being subject to a *ceteris paribus* (other things being equal) clause.[79] It is a considerable weakness, because "a weak prediction, that the event will occur, if there is no disturbances, cannot be confirmed, nor can it be refuted, by what is found to happen."[80]

So, if Hick's idea of the weakness of economic prediction is correct, then economics is not a science or, at least, not yet. Science needs *conditional strong predictions*, where the predicted event will occur, under stated conditions, and the prediction will be either confirmed or refuted by the event. Otherwise, if prediction is so weak —as he holds— such that its success or failure cannot be assessed by the experience, then we are at the boundaries of non-scientific prediction. In my judgment, Hicks goes too far when he maintains these views, because in Economics we can find foresights.[81] A *foresight* takes place when there is a clear control of the variables involved: it shows the state of a variable within a period of time, when that variable is directly or indirectly under our control (e.g., the interest rates on the immediate horizon; the collection of some kind of taxes in the short run, such as the VAT of some products; the number of civil servants next year in a nation where this matter is regulated by the government…).

In other words, economic predictions are not *eo ipso* essentially weak predictions. Although it is clear that the *ceteris paribus* should be considered,[82] the internal and external limits do not imply such *essential* weakness; rather, they suggest that knowledge in economics is more difficult than in physics (or natural sciences, in general) and is a more complex reality than the physical one (in terms of interrelation of factors of economic

[78] Hicks, J., "Is Economics a Science?" p. 94.

[79] Cfr. "Is Economics a Science?" p. 97.

[80] Hicks, J., "Is Economics a Science?" p. 94. "If the event does occur, all we can say is that if the prediction was correct, there appear to have been no disturbances, or no sufficient disturbances. If the event does not occur, we can say no more than that if the prediction was correct, then there were disturbances. In neither case is it shown, at all directly, by experience that the prediction was right or wrong", Hicks, J., *Ibidem*.

[81] Cf. Fernandez Valbuena, S., "Predicción y Economía," in Gonzalez, W. J. (ed.), *Aspectos metodológicos de la investigación científica*, 2nd ed., Ediciones Universidad Autónoma de Madrid and Publicaciones Universidad de Murcia, Madrid-Murcia, 1990, pp. 385-405.

[82] Cf. Boumans, M. and Morgan, M. S., "*Ceteris paribus* Conditions: Materiality and the Application of Economic Theories," *Journal of Economic Methodology*, v. 8, n. 1, (2001), pp. 11-26.

activity and as a consequence of the connections with many other human activities). Therefore, it is a different kind of science: a human and social one which can also be developed as a science of the artificial (especially, in the sector of "financial products"); and the model of Newtonian Mechanics does not seem to be valid for Economics. The phenomena studied by Economics are in the range of being influenced by human action and contain conditions that the economics could incorporate into predicting models.[83]

3.2.3. Prediction Split: Buchanan's Dualist Posture

Buchanan's *methodological dualism* also goes too far. His proposal wants to show that economics has to study choices of economic agents and, at the same time, that to predict results of those interactions can be problematic or even unreliable.[84] His epistemological basis is dualist as well: subjective knowledge versus objective knowledge, which is predictive. However, subject matter of economics has to be *objective*, otherwise it cannot be "science" in a strict sense. Hence, although the existence of subjective factors cannot be denied (they have been studied by Kahneman, among others), economic models should be able to integrate the subjective side of economic activity. It cannot be excluded a priori from the scientific realm that studies economic reality. Moreover, economic models should deal with variables of the subjective side of economic performance[85] and to have characterizations, such as *bounded rationality*,[86] that fit much better with the activities of economic agents than other visions of mainstream economics (among them, rationality as maximization).

Nonetheless, Buchanan is opposed to another extreme, defended by some specialists that see economics as purely "objective" discipline, in the sense of being akin to a branch of mathematics. Understood in this way, economics appears as the intersection of pure axiomatization and applied geometry, which can be focused on formal properties (e.g., either of a set of assumptions about the transitivity of abstract relations in the case of the notion of "rationality," or of the results of general equilibrium theory). For Rosenberg, "if this view is correct we cannot demand that it provide the reliable guide to the behavior of economic agents and the performance of economics as a whole for which the formulation of public policy looks to economics. We should neither attach much confidence to predictions made on its basis nor condemn it severely when these predictions fail."[87]

---

83 Statistical models are of different kinds, according to the variables involved: "A major use of these models has been to provide short and medium term forecasts for important macro variables, such as consumption, income, investment, and unemployment, all of which are integrated series. The derived growth rates are found to be somewhat forecastable. Much less forecastable are inflation rates and returns from speculative markets, such as stocks, bonds, and exchange rates," GRANGER, C. W. J., "Time Series Analysis, Cointegration, and Applications," in FRÄNGSMYR, T. (ed.), *From Les Prix Nobel. The Nobel Prizes 2003*, Nobel Foundation, Stockholm, 2004, p. 364.

84 Cf. BUCHANAN, J. M., "Is Economics the Science of Choice?" in BUCHANAN, J. M., *Economics: Between Predictive Science and Moral Philosophy*, pp. 35-50.

85 This facet could be sometimes "subjectual" (i.e., objective in the individuals). Subjectual belongs to the subject, but it is not individualistic (i.e., idiosyncratic) or purely individual (e.g., cognitive limitations in the decision-making appear in each agent, but it is "subjectual" insofar as this feature happens objectively in every individual; the subjective part is the level of intensity or the variations according to the environment).

86 Cf. SIMON, H. A., "Bounded Rationality in Social Science: Today and Tomorrow," *Mind and Society*, v. 1, n. 1, (2000), pp. 25-39.

87 ROSENBERG, A., "If Economics isn't Science, What is It?," in HAUSMAN, D. M. (ed.), *The Philosophy of Economics. An Anthology*, 2nd ed., revised, Cambridge University Press, Cambridge, 1994, p. 391.

Taking into account the analysis of the positions on economic prediction of three Nobel Prize laureate (Friedman, Hicks, and Buchanan), the binomial relation *acceptance of prediction-guarantee of scientific character* of economics should be emphasized. From different perspectives, these three economists connect reliable prediction to the fact of having science; and, the other way round, when they consider that there is no reliable economic predictions, the theoretical consequence is clear: that the scientific status of economics disappears. In other words, prediction works *de facto* as test of scientific character of economics. Furthermore, as Hutchison has pointed out, the main questions of interest for economics are around the topic of predictions.[88] Thus, the question 'are there scientific predictions in economics?', which receives these responses (affirmative, negative, and partially affirmative and negative), is central for this discipline.

In spite of their differences, these three positions (predictivist, quasi-scientific, and dualist) endorse the idea of prediction as an *aim* of economics, but they disagree about its role as the scientific *test*. Two of them agree on the existence of a realm —positive economics— where this discipline is an objective science, capable of predicting future economic phenomena in a reliable way. This can be reinterpreted in terms of predictability of "economic activity", i.e., when the variables which are studied can be separated or distinguished ("isolated") with respect to other activities, or when they can be seen in an ideal context (such as some theorems of economic theory and also econometric theorems). In this case, the economic model may be precise, like a mere mathematical calculation, but it may put aside the usual variables that interfere in the real case. From this point of view, prediction (or, even, foresight) of future phenomena can turn out to be viable, because it is possible to take into account the variable involved in the event and, thus, have a precise knowledge of the outcome.

3.2.4. Prediction interconnected with Uncertainty and Bounded Rationality: Simon's Scheme

Although Herbert Simon avoids more acute problems of Friedman's methodological predictivism,[89] and he has no doubt about the scientific character of economics, he does not establish clear boundaries between scientific predictions and non-scientific ones. Nonetheless, his "empirically grounded economic reason" gives a more realistic picture of economic behavior than the previous approaches, especially in the field of microeconomics. Furthermore, his philosophic-methodological conception establishes links between uncertainty, bounded rationality and prediction. Thus, his *scheme* presents several elements in his economic theory, which are all interconnected.[90]

---

[88] "The question of prediction in Economics involves, or brings together, most of the main questions as to what sort of subject Economics is," HUTCHISON, T. W., "On Prediction and Economic Knowledge," in HUTCHISON, T. W., *Knowledge and Ignorance in Economics*, p. 8.

[89] "The question whether a theory is realistic 'enough' can be settled only by seeing whether it yields to predictions that are good enough for the purpose at hand or that are better than predictions from alternative theories," FRIEDMAN, M., "The Methodology of Positive Economics", p. 41.

[90] On the relation between rationality and prediction in Simon, cf. GONZALEZ, W. J., "Racionalidad y Economía: De la racionalidad de la Economía como Ciencia a la racionalidad de los agentes económicos," in GONZALEZ, W. J. (ed.), *Racionalidad, historicidad y predicción en Herbert A. Simon*, Netbiblo, A Coruña, 2003, pp. 65-96; especially, pp. 83-86.

i) Economic prediction cannot be made on the basis of a "perfect rationality." There are *limitations* for predictions in economics: the uncertainty about the consequences that would follow from each alternative when a decision is made, the information about the set of alternatives could be incomplete, and the complexity of the situations may prevent us from making the necessary computations to solve the problem.[91] In other words, human cognitive boundaries raise serious difficulties as to the predictive success in economics. (His focus is on microeconomics but it can be enlarged to cover macroeconomic events as well.)

ii) Insofar as the consequences of human behavior are extended into the future, we need *correct predictions* for objectively rational choices. But, on the basis of realism of assumptions, prediction is not a pure inference on the grounds of optimal conditions,[92] because human decision-making is rooted in procedures that lead to the strategy of *satisficing* rather than to optimizing.[93] Hence, for Simon, in order to make an appropriate economic prediction, we need to know about some things that belong to the natural environment (e.g., the weather), others related to the social and political environments beyond the economic (e.g., a revolution or a civil war), in addition to the behavior of other economic actors (customers, competitors, suppliers,) which may influence our own behaviors.[94]

iii) Even though there is a clear recognition by Simon that his approach to economic predictions also involves limitations, he is convinced that his vision of economic prediction based on a bounded rationality framework *fits* with a lot of things.[95] His conviction connects with the epistemological grounding: he offers a viewpoint that seeks the reality of human behavior as it can be observed in economic life.[96] Thus, he considers his position more realistic than the conception of mainstream economics.

iv) Unexpected events should be taken into account as well. Hence, in addition to the estimation of the probabilities of predicted events, we need to use *feedback* to correct the unexpected phenomena. Simon acknowledges the need for a revision mechanism to diminish the number of errors: "a system can generally be steered more accurately if it uses feedforward, based on predictions of the future, in combination with feedback, to correct the errors of the past. However, forming expectations to deal with uncertainty creates its own problems."[97]

---

[91] Cf. SIMON, H. A., "Theories of Bounded Rationality," in MCGUIRE, C. B. and RADNER, R. (eds.), *Decision and Organization*, North-Holland, Amsterdam, 1972, p. 169.

[92] "Human beings (and other creatures) do not behave optimally for their fitness, because they are wholly incapable of acquiring the knowledge and making the calculations that support optimization," SIMON, H. A., "Altruism and Economics," *American Economic Review*, v. 83, n. 2, (1993), p. 156.

[93] His vision of the future of bounded rationality is rooted in that idea, cf. SIMON, H. A., "Bounded Rationality in Social Science: Today and Tomorrow," pp. 25-39.

[94] Cf. *The Sciences of the Artificial*, 3rd ed., p. 35.

[95] Cf. SIMON, H. A., "Colloquium with H. A. Simon," in EGIDI, M. and MARRIS, R. (eds.), *Economics, Bounded Rationality and the Cognitive Revolution*, p. 18.

[96] Cf. SIMON, H. A., "Introductory Comment," p. 3. "Because game theory is intrinsically unable to make specific predictions of behaviour from the postulates of rationality, in order to understand behaviour we must look at the empirical phenomena to see how people actually play games," SIMON, H. A., "Colloquium with H. A. Simon," p. 25.

[97] *The Sciences of the Artificial*, 3rd ed., p. 36.

If this epistemological-methodological scheme, which interconnects some key elements —uncertainty, bounded rationality, and prediction— is compared with the approach of mainstream economics (subjective expected utility of neoclassical economics), the scales tip in favor of Simon in four aspects. Nevertheless, his position is not good enough to resolve the problem, insofar as the focus of the analysis of predictions in economics should be economic "activity" rather than the case of "behavior."[98]

a) Certainly there is an improvement with respect to the *realism of assumptions*, because Simon pays attention to economic reality as it is, he avoids drawing strong conclusions from a few *a priori* assumptions.[99] The emphasis on uncertainty and bounded rationality is quite pertinent to the problem of economic predictions, in addition to the insistence that the attempt to predict human economic behavior by deductive inference from a small set of unchallengeable premises is a failure.[100]
b) Clearly, the nexus between economics and psychology makes sense not only for the case of rational choice but also in predicting the results of interactions of economic actors,[101] because —in my judgment— economics is based on a *human activity*. The explanation and prediction of the movements of the economy are related to the actions of the members of society and to the interactions of their activities.[102] In other words, the elements of "economic activity" and "economics as activity" have direct implications for the realm of prediction.[103] On the one hand, the normal aim of a human activity is more connected with present circumstances than with a future not yet observed. On the other hand, the predictability of economic activity —which is, in principle, autonomous— is possible, and could be reliable; whereas predictability of economics as a human activity among others appears more unreliable, precisely due to the interdependence with other activities.
c) To be sure, prediction has a crucial role in economics: the development of economic activities requires the anticipation and, if it is possible, the control of future events. However, *pace* Friedman, not all of economics is concerned with predicting. In fact, Simon has stressed the need of *prescription* in systems modeling,[104] which directly affects the characterization of economics. Moreover, his vision of economics as applied science led him to give more methodological weight to the task of prescription than to prediction.

---

98 This alternative is developed in GONZALEZ, W. J., "Rationality in Economics and Scientific Predictions: A Critical Reconstruction of Bounded Rationality and its Role in Economic Predictions," pp. 222-229. The content of this conception was discussed personally with Herbert Simon on 15 August 1996.

99 The realism of assumptions has also been defended by other authors, especially Paul Samuelson in his direct criticisms of Friedman's methodology, cf. SAMUELSON, P., "Problems of Methodology-Discussion," *American Economic Review*, v. 53, n. 2, (1963), pp. 231-236.

100 Cf. SIMON, H. A., "From Substantive to Procedural Rationality," in LATSIS, S. (ed), *Method and Appraisal in Economics*, Cambridge University Press, Cambridge, 1976, p. 146.

101 Cf. SIMON, H. A., "Economics and Psychology," in KOCH, S. (ed.), *Psychology: A Study of a Science*, vol. 6, pp. 715-752. Reprinted in SIMON, H. A., *Models of Bounded Rationality, vol. 2: Behavioral Economics and Business Organization*, pp. 318-355.

102 Cf. SEN, A., "Prediction and Economic Theory," p. 14.

103 Cf. GONZALEZ, W. J., "Economic Prediction and Human Activity. An Analysis of Prediction in Economics from Action Theory," pp. 253-294; especially, pp. 262-280.

104 Cf. SIMON, H. A., "Prediction and Prescription in Systems Modeling," pp. 7-14.

Thus, Simon seeks to *prescribe* human economic behavior, but not by means of a deductive inference from a small set of premises. Prediction and prescription are interconnected: "predictive models are only a special case where we seek to predict events we cannot control in order to adapt to them better. We do not expect to change the weather, but we can take steps to moderate its effects. We predict populations so that we can plan to meet their needs for food and education. We predict business cycles in order to plan our investment and production levels. (…) When our goal is prescription rather than prediction, then we can no longer take it for granted that what we want to compute are time series."[105]

d) Simon has rightly rejected the primacy of methodological instrumentalism based on predictive success. For him, the crucial point for the assessment of economic predictions is the reliability of the *method* used to make predictions rather than the accuracy of the result. Thus, prediction based on the bounded rationality model avoids central problems of the instrumentalist position (held, among others, by Friedman).[106] However, the model based on economics understood as a human activity reflects better the complexity of economic predictions and the necessity of combining the rationality of ends (goals) with the rationality of means (procedures), because Simon only works on the basis of the latter —instrumental rationality— and skips the former (evaluative rationality).[107]

4. Coda on Prediction as Scientific Test

Besides the problem of the role of prediction as a test of economic theory, there is a line of research in econometrics which not only assumes the *need* to make predictions in order to have the science of economics but also stresses that aspect. Thus, authors as Clive W. J. Granger, who makes a test of causality that is the base of economic predictions in multivariate models; Thomas Sargent who with Robert E. Lucas Jr. (Nobel Prize winner in 1995) makes the test of contrast of rational expectations; David F. Hendry, who constructs the major part of the model of dynamic prediction in macroeconomics; and Andrew C. Harvey, who presents the problems which arise from the specification of models to make predictions.[108]

Nevertheless, these studies can sometimes work on "idealized" settings or a certain kind of projections of time-series based on data of the past. But economics, as a real activity that is connected with many other activities, usually has more problems than those merely derived from variables of pure "economic activity," and that has a repercussion in

[105] Simon, H. A., "Prediction and Prescription in Systems Modeling," pp. 10-11.

[106] On instrumentalism about scientific prediction, cfr. Gonzalez, W. J., "Reichenbach's Concept of Prediction," pp. 35-56; especially, pp. 43-50.

[107] Cf. Simon, H. A., *Reason in Human Affairs*, Stanford University Press, Stanford, 1983, pp. 7-8

[108] This synthesis on prediction and econometrics was suggested by Carlos Fernández-Jardón, *Personal Communication*, 4 August 2005. A historical background is available in Morgan, M. S., *The History of Econometric Ideas*, Cambridge University Press, Cambridge, 1990; Qin, D., *The Formation of Econometrics*, Clarendon Press, Oxford, 1993: and Epstein, R. J., *A History of Econometrics*, North Holland, Amsterdam, 1987. On the methodological aspects, cf. Poirier, D. J. (ed), *The Methodology of Econometrics*, Edward Elgar, Aldershot, 1994, 2 vol.

giving an accurate prediction.[109] Moreover, the possibility itself of prediction with rigor can be questioned. The variables that are under our control could be affected by other variables that are not controlled by us, either because it is not possible at this moment to know of their existence, or because of the (temporary) impossibility of getting information about them.

When this second possibility appears, it is said to be an "unpredictable" phenomenon. In that case, economics as activity could have —at least, theoretically— literally unpredictable events. However, in spite of this possibility, it seems more adequate to call these phenomena "unexpected" or "not predictable," because if a phenomenon could be explained afterwards, then it could have been predicted previously (for example, something extremely rare), even though still is "not predictable" with the present means.[110]

Economics has, then, two kinds of problems in this field: on the one hand, those derived from the *present situation* of our knowledge about economic affairs (i.e., the insufficient knowledge of a "knowable" reality); and, on the other hand, the intrinsic *mutability* of the economic reality, mainly when this is within the real sphere of economics as an activity interdependent of others. There can be seen the defective character of economic knowledge —its insufficient domain— due, precisely, to the complexity of economic reality, which interacts with other human activities (social, cultural, political, ecological, ...).

Neither problem is exclusive of prediction in economics, because —in some way or another— they are present in other human and social sciences as well. Behind these problems is something basic which is frequently forgotten in economic predictions (and forecasting): the human and social character of economic activity and its interdependence on other human activities. So, as something *human* developing in the social environment, it cannot have the *same* characteristics as the natural phenomena. In other words, the continuous efforts to develop economic predictions, as a mere calculation, are not well focused due to the existence of human factors (qualitative ones), which cannot be expressed in reducible quantitative terms. Hence, there are limitations for economic models, as Clive Granger has clearly recognized.[111]

Through the development of statistical economics and econometrics, where probability calculus (classical and Bayesian) is crucial, there is a methodological convergence with physics. Economic models and physical models rely on probability calculus for predicting. To some extent, they share the problems of accuracy and precision of predictions, depending on the kind of event that they predict. There are also some physical phenomena whose predictability is questioned, even in determinist contexts. In this regard, the theory

---

[109] Robert Solow, Nobel Prize winner in 1987, has pointed out that "*much of the most interesting macroeconomic theory being done today* is an attempt to investigate coordination failures in the economy. They are much like external effects, or what in the U. S. academic circles are called thin market externalities," in SOLOW, R., *The Rate of Return and the Rate of Interest*, The Industrial Institute for Economic and Social Research, Stockholm, 1991, p. 25.

[110] I owe to Patrick Suppes this distinction between "unpredictable" and "not predictable." *Personal Communication*, November 1993.

[111] "The modern macro economy is large, diffuse, and difficult to define, measure, and control. Economists attempt to build models that will approximate it, that will have similar major properties so that one can conduct simple experiments on them, such as determining the impacts of alternative policies or the long-run implications of some new institution," GRANGER, C. W. J., "Time Series Analysis, Cointegration, and Applications," in FRÄNGSMYR, T. (ed.), *From Les Prix Nobel. The Nobel Prizes 2003*, p. 362.

of chaos focuses on the predictability of diverse physical events and gives reasons why the prediction of certain phenomena may be impossible.[112]

To sum up: the problem of prediction as scientific test of economics, in general, and economic theory, in particular, can be resolved assuming that *accuracy* and *precision* of prediction serve as *sufficient condition* for having science in the case of economics, but it is not a necessary condition for its scientific status. Predictive success is a weak frontier for demarcation, insofar as economic theories (in microeconomics and in macroeconomics) can have other criteria of scientific validity (e.g., in terms of a scientific explanation that achieves enough evidence). Economics cannot be subordinated to the aim of prediction, even though the reliability of economic predictions could be very important, both to evaluate the quality of an economic theory and its repercussion for prescriptions of applied economics.

Finally, an extensive *philosophic-methodological framework* for prediction in economics should include several aspects. 1) Semantically, "prediction" is a statement about the future: the sense of anticipation could be ontological or epistemological, and the reference may be a new fact and some non-observed phenomena.[113] 2) Logically, scientific prediction is structurally different from scientific explanation, insofar as an explanation is an argument (and thus is focused on the answer to the question why?). 3) Epistemologically, prediction requires an empirical grounding, and scientific theories can offer a reasonable base for practical prediction.[114] Therefore, an underlying rationality is always needed to every scientific prediction made.

4) Methodologically, it is still an open discussion on whether prediction has more weight on confirmation than accommodation to known facts. Nevertheless, mainstream economics (principalyly in economic theory but also in econometrics) emphasizes customarily prediction over accommodation. 5) Ontologically, not all economic activity —nor economics as activity— should be oriented towards prediction, because understanding economic

[112] On the theory of chaos, cf. Batterman, R. W., "Defining Chaos," *Philosophy of Science*, v. 60, n. 1, (1993), pp. 43-66; Cvitanovic, P. (ed.), *Universality in Chaos*, Hilger, Bristol, 1984; Kellert, S. E., *In the Wake of Chaos. Unpredictable Order in Dynamical Systems*, The University of Chicago Press, Chicago, 1993; Peitgen, H. O., Jürgens, H. and Saupe, D., *Chaos and Fractals. New Frontiers of Science*, Springer, N. York, 1992; Ruelle, D., *Chance and Chaos*, Princeton University Press, Princeton, 1991; and Winnie, J. A., "Computable Chaos," *Philosophy of Science*, v. 59, n. 2, (1992), pp. 263-275. For Herbert A. Simon, we "don't know whether the economy is a chaotic system," in Sichel, W. (ed.), *The State of Economic Science. Views of Six Nobel Laureates*, p. 99. He maintains that "chaos is essentially a statistical condition. It does not imply that *anything goes*. We may, for example, despair of tracing the future course of business cycles without renouncing the goal of making statements about the long-run development of an economy. For instance, we might make perfectly sound statements about the upper limits on per capita GNP without being able to say how closely, or when, these limits will be approached," Simon, H. A., "Prediction and Prescription in Systems Modeling," p. 8.

[113] Ontologically, the reality itself which is predicted do not require to have *eo ipso* posterior existence to the predictive statement, because it is legitimated to state in advance a social or economical event which, strictly speaking, is already going on (as was the case in astronomy with the prediction of Neptune or in quantum mechanics with the existence of neutrino). In human contexts, when a person is qualified as "predictable", he or she could be seen as a "reliable person," and that means that the person is well-known.

[114] "Typically there will be an infinite array of generalisations which are compatible with the available observational evidence, and which are therefore, as yet, unrefuted. If we were free to choose arbitrarily from among all the unrefuted alternatives, we could predict anything whatever. If there were no rational basis for choosing from among all the unrefuted alternatives, then, as I think Popper would agree, there would be no such thing as rational prediction," Salmon, W. C., "Rational Prediction," *The British Journal for the Philosophy of Science*, v. 32, (1981), p. 117.

phenomena is important in different kinds of processes, such as decision-making in business firms. 6) Axiologically, the relevance of prediction as economic aims is great, especially when it is accompanied by values such as accuracy and precision.

5. Bibliography

Backhouse, R. E. (ed.), *New Directions in Economic Methodology*, Routledge, London, 1994.

Backhouse, R. E., *Truth and Progress in Economic Knowledge*, E. Elgar, Cheltenham, 1997.

Batterman, R. W., "Defining Chaos," *Philosophy of Science*, v. 60, n. 1, (1993), pp. 43-66.

Blaug, M., *The Methodology of Economics: Or How Economists Explain*, Cambridge University Press, Cambridge, 1980.

Blaug, M., "John Hicks and the Metodology of Economics," in Marchi, N. de (ed.), *The Popperian Legacy in Economics*, Cambridge University Press, Cambridge, 1988, pp. 183-195.

Boumans, M. and Morgan, M. S., "*Ceteris paribus* Conditions: Materiality and the Application of Economic Theories," *Journal of Economic Methodology*, v. 8, n. 1, (2001), pp. 11-26.

Boylan, Th. A. and O'Gorman, P. F., *Beyond Rhetoric and Realism in Economics. Towards a Reformulation of Economic Methodology*, Routledge, London, 1995.

Buchanan, J. M., *Economics: Between Predictive Science and Moral Philosophy*, Texas A & M University Press, College Station, 1987.

Buchanan, J. M., "Positive Economics, Welfare Economics, and Political Economy," in Buchanan, J. M., *Economics: Between Predictive Science and Moral Philosophy*, Texas A & M University Press, College Station, 1987, pp. 3-19.

Buchanan, J. M., "What Should Economists Do?," in Buchanan, J. M., *Economics: Between Predictive Science and Moral Philosophy*, Texas A & M University Press, College Station, 1987, pp. 21-33.

Buchanan, J. M., "Is Economics the Science of Choice?," in Buchanan, J. M., *Economics: Between Predictive Science and Moral Philosophy*, Texas A & M University Press, College Station, 1987, pp. 35-50.

Buchanan, J. M., "The Domain of Subjective Economics: Between Predictive Science and Moral Philosophy," in Buchanan, J. M., *Economics: Between Predictive Science and Moral Philosophy*, Texas A & M University Press, College Station, 1987, pp. 67-80.

Buchanan, J. M., "The State of Economic Science," in Sichel, W. (ed.), *The State of Economic Science. Views of Six Noble Laureates*, W. E. Upjohn Institute for Employment Research, Kalamazoo, Michigan, 1989, pp. 79-95.

Buchanan, J. M., "Economics as a Public Science," in Medema, S. G. and Samuels, W. J. (eds.), *Foundations of Research in Economics: How do Economists do Research*, E. Elgar, Brookfield, VT, 1996, pp. 30-36.

Caldwell, B., *Beyond Positivism: Economic Methodology in the Twentieth Century*, Allen and Unwin, London, 1982; revised edition, Routledge, London, 1994.

Carnap, R., "Theory and Prediction in Science," *Science*, v. 104, n. 2710, (6 December 1946), pp. 520-521.

CVITANOVIC, P. (ed.), *Universality in Chaos*, Hilger, Bristol, 1984.

EPSTEIN, R. J., *A History of Econometrics*, North Holland, Amsterdam, 1987.

FERNANDEZ VALBUENA, S., "Predicción y Economía," in GONZALEZ, W. J. (ed.), *Aspectos metodológicos de la investigación científica*, 2nd ed., Ediciones Universidad Autónoma de Madrid and Publicaciones Universidad de Murcia, Madrid and Murcia, 1990, pp. 385-405.

FRIEDMAN, M., "The Methodology of Positive Economics," in FRIEDMAN, M., *Essays in Positive Economics*, The University of Chicago Press, Chicago, 1953 (6th reprint, 1969), pp. 3-43.

GONZALEZ, W. J., "Economic Prediction and Human Activity. An Analysis of Prediction in Economics from Action Theory," *Epistemologia*, v. 17, (1994), pp. 253-294.

GONZALEZ, W. J., "Reichenbach's Concept of Prediction," *International Studies in the Philosophy of Science*, v. 9, n. 1, (1995), pp. 35-56.

GONZALEZ, W. J., "On the Theoretical Basis of Prediction in Economics," *Journal of Social Philosophy*, v. 27, n. 3, (1996), pp. 201-228.

GONZALEZ, W. J., "Rationality in Economics and Scientific Predictions: A Critical Reconstruction of Bounded Rationality and its Role in Economic Predictions," *Poznan Studies in the Philosophy of Science*, v. 61, (1997), pp. 205-232.

GONZALEZ, W. J., "Prediction and Prescription in Economics: A Philosophical and Methodological Approach," *Theoria*, v. 13, n. 32, (1998), pp. 321-345.

GONZALEZ, W. J., "Marco teórico, trayectoria y situación actual de la Filosofía y Metodología de la Economía," *Argumentos de la Razón Técnica*, v. 3, (2000), pp. 13-59.

GONZALEZ, W. J., "Lakatos's Approach on Prediction and Novel Facts," *Theoria*, v. 16, n. 3, (2001), pp. 499-518.

GONZALEZ, W. J., "From *Erklären-Verstehen* to *Prediction-Understanding*: The Methodological Framework in Economics," in SINTONEN, M., YLIKOSKI, P. and MILLER, K. (eds.), *Realism in Action*: *Essays in the Philosophy of Social Sciences*, Kluwer, Dordrecht, 2003, pp. 33-50.

GONZALEZ, W. J., "Racionalidad y Economía: De la racionalidad de la Economía como Ciencia a la racionalidad de los agentes económicos," in GONZALEZ, W. J. (ed.), *Racionalidad, historicidad y predicción en Herbert A. Simon*, Netbiblo, A Coruña, 2003, pp. 65-96.

GONZALEZ, W. J., "The Many Faces of Popper's Methodological Approach to Prediction," in CATTON, PH. and MACDONALD, G. (eds.), *Karl Popper: Critical Appraisals*, Routledge, London, 2004, pp. 78-98.

GONZALEZ, W. J., "The Philosophical Approach to Science, Technology and Society," in GONZALEZ, W. J. (ed.), *Science, Technology and Society: A Philosophical Perspective*, Netbiblo, A Coruña, 2005, pp. 3-49.

GRANGER, C. W. J. and NEWBOLD, P., *Forecasting Economic Time Series*, Academic Press, N. York, 1977.

GRANGER, C. W. J., *Forecasting in Business and Economics*, 2nd ed., Academic Press, S. Diego, 1989 (1st ed., 1980).

GRANGER, C. W. J., "Where are the Controversies in Econometric Methodology?," in GRANGER, C. W. J. (ed.), *Modelling Economics Series: Readings in Econometric Methodology*, Clarendon Press, Oxford, 1990, pp. 1-23.

GRANGER, C. W. J. and PESARAN, M. H., "Economic and Statistical measures of Forecast Accuracy," *Journal of Forecasting*, v. 19, (2000), pp. 537-560.

GRANGER, C. W. J., "Evaluation of Forecasts," in HENDRY, D. F. and ERICSSON, N. R. (eds.), *Understanding Economic Forecasts*, The MIT Press, Cambridge (MA), 2002, pp. 93-103.

GRANGER, C. W. J. and POON, S., "Forecasting Volatility in Financial Markets," *Journal of Economic Literature*, v. 41, (2003), pp. 478-539.

GRANGER, C. W. J., "Time Series Analysis, Cointegration, and Applications," in FRÄNGSMYR, T. (ed.), *From Les Prix Nobel. The Nobel Prizes 2003*, Nobel Foundation, Stockholm, 2004, pp. 360-366.

HANDS, D. WADE, *Reflection without Rules. Economic Methodology and Contemporary Science Theory*, Cambridge University Press, Cambridge, 2001.

HAUSMAN, D. M., *The Inexact and Separate Science of Economics*, Cambridge University Press, Cambridge, 1992.

HICKS, J., "A Discipline not a Science," in HICKS, J., *Classics and Moderns. Collected Essays on Economic Theory, v. III*, Harvard University Press, Cambridge, 1983, pp. 364-375.

HICKS, J., "Is Economics a Science?," in BARANZINI, M. and SCAZZIERI, R. (eds.), *Foundations of Economics. Structures of Inquiry and Economic Theory*, B. Blackwell, Oxford, 1986, pp. 91-101.

HUTCHISON, T. W., *"Positive" Economics and Policy Objectives*, Allen and Unwin, London, 1964.

HUTCHISON, T. W., "On Prediction and Economic Knowledge," in HUTCHISON, T. W., *Knowledge and Ignorance in Economics*, Blackwell, Oxford, 1977, pp. 8-33 and 145-151.

HUTCHISON, T. W., "To Predict or not to Predict? (That *is* the Question)," in HUTCHISON, T. W., *Changing Aims in Economics*, B. Blackwell, Oxford, 1992, pp. 71-88 and 158-167 (notas).

KAHNEMAN, D. and TVERSKY, A., "On the Psychology of Prediction," *Psychological Review*, v. 80, (1973), pp. 237-251.

KAHNEMAN, D. and TVERSKY, A., "Prospect Theory: An Analysis of Decisions Under Risk," *Econometrica*, v. 47, (1979), pp. 313-327.

KAHNEMAN, D. and SNELL, J., "Predicting Utility," in HOGARTH, R. M. (ed.), *Insights in Decision Making*, The University of Chicago Press, Chicago, 1990, pp. 295-310.

KAHNEMAN, D., "Maps of Bounded Rationality: Psychology for Behavioral Economics," *The American Economc Review*, v. 93, n. 5, (2003), pp. 1449-1475.

KELLERT, S. E., *In the Wake of Chaos. Unpredictable Order in Dynamical Systems*, The University of Chicago Press, Chicago, 1993.

KITCHER, PH., *Science, Truth and Democracy*, Oxford University Press, N. York, 2001.

KUHN, TH. S., "Postscript-1969," in KUHN, TH. S., *The Structure of Scientific Revolutions*, The University of Chicago Press, Chicago, 2nd ed., 1970, pp. 174-210.

LAUDAN, L., *Progress and its Problems*, University of California Press, Berkeley, 1977.

LAUDAN, L., "Una Teoría de la evaluación comparativa de teorías científicas," in GONZALEZ, W. J. (ed.), *El Pensamiento de L. Laudan. Relaciones entre Historia de la Ciencia y Filosofía de la Ciencia*, Publicaciones de la Universidad de A Coruña, A Coruña, 1998, pp. 155-169.

MATHIES, B. P. and DIAMANTIPOULOS, A., "Towards a Taxonomy of Forecast Error Measures. A Factor-comparative Investigation of Forecast Error Dimensions," *Journal of Forecasting*, v. 13, (1994), pp. 409-416.

MORGAN, M. S., *The History of Econometric Ideas*, Cambridge University Press, Cambridge, 1990.

MCNEES, S. K., "Why do Forecasts Differ?," *New England Economic Review*, January-February, (1989), pp. 42-52.

NIINILUOTO, I., "The Aim and Structure of Applied Research," *Erkenntnis*, v. 38, (1993), pp. 1-21.

PEITGEN, H. O., JÜRGENS, H. and SAUPE, D., *Chaos and Fractals. New Frontiers of Science*, Springer, N. York, 1992.

POIRIER, D. J. (ed.), *The Methodology of Econometrics*, Edward Elgar, Aldershot, 1994, 2 vol.

POPPER, K. R., *Logik der Forschung*, Julius Springer Verlag, Vienna, 1935. Translated by the author: *The Logic of Scientific Discovery*, Hutchinson, London, 1959.

POPPER, K., *The Poverty of Historicism*, Routledge and K. Paul, London, 1957.

QIN, D., *The Formation of Econometrics*, Clarendon Press, Oxford, 1993.

RADNITZKY, G., "The Boundaries of Science and Technology," in: *The Search for Absolute Values in a Changing World. Proceedings of the Sixth International Conference on the Unity of the Sciences. Vol. 2*, International Cultural Foundation Press, N. York, 1978, pp. 1007-1036.

REICHENBACH, H., *Experience and Prediction*, The University of Chicago Press, Chicago, 1938.

RESCHER, N., *Predicting the Future*, State University Press New York, N. York, 1998.

RESCHER, N., *The Limits of Science*, revised edition, University of Pittsburgh Press, Pittsburgh, 1999.

ROSENBERG, A., *Economics-Mathematical Politics or Science of Diminishing Returns*?, The University of Chicago Press, Chicago, 1992.

ROSENBERG, A., "If Economics isn't Science, What is It?," in HAUSMAN, D. M. (ed.), *The Philosophy of Economics. An Anthology*, 2nd ed., revised, Cambridge University Press, Cambridge, 1994, pp. 376-394.

ROSENBERG, A., "La Teoría Económica como Filosofía Política," *Theoria*, v. 13, n. 32, (1998), pp. 279-299.

RUELLE, D., *Chance and Chaos*, Princeton University Press, Princeton, 1991.

SALMON, W. C., "Rational Prediction," *The British Journal for the Philosophy of Science*, v. 32, (1981), pp. 115-125.

SALMON, W. C., "On the Alleged Temporal Anisotropy of Explanation," in EARMAN, J., JANIS, A., MASSEY, G. and RESCHER, N. (eds.), *Philosophical Problems of the Internal and External Worlds*, University of Pittsburgh Press, Pittsburgh, 1993, pp. 229-248.

Samuelson, P., “Problems of Methodology-Discussion,” *American Economic Review*, v. 53, n. 2, (1963), pp. 231-236.

Sen, A., “Prediction and Economic Theory,” in Mason, J., Mathias, P. and Westcott, J. H. (eds.), *Predictability in Science and Society*, The Royal Society and The British Academy, London, 1986, pp. 103-125.

Simon, H. A., “Economics and Psychology,” in Koch, S. (ed.), *Psychology: A Study of a Science*, vol. 6, McGraw Hill, N. York, 1963, pp. 715-752. Reprinted in Simon, H. A., *Models of Bounded Rationality, vol. 2: Behavioral Economics and Business Organization*, The MIT Press, Cambridge, 1982, pp. 318-355.

Simon, H. A., “Theories of Bounded Rationality,” in McGuire, C. B. and Radner, R. (eds.), *Decision and Organization*, North-Holland, Amsterdam, 1972, pp. 161-176.

Simon, H. A., “From Substantive to Procedural Rationality,” in Latsis, S. (ed.), *Method and Appraisal in Economics*, Cambridge University Press, Cambridge, 1976, pp. 129-148.

Simon, H. A., “Rational Decision Making in Business Organizations,” *American Economic Review*, v. 69, n. 4, (1979), pp. 493-513.

Simon, H. A., *Reason in Human Affairs*, Stanford University Press, Stanford, 1983.

Simon, H. A., “Testability and Approximation,” in Hausman, D. M. (ed.), *Philosophy of Economics*, Cambridge University Press, Cambridge, 1984 (1st ed.), pp. 245-248. Originally was published as Simon, H. A., “Problems of Methodology-Discussion,” *American Economic Review*, v. 53, (1963), pp. 229-231.

Simon, H. A., “Rationality in Psychology and Economics,” in Hogarth, R. M. and Reder, M. W. (eds.), *Rational Choice. The Contrast between Economics and Psychology*, The University of Chicago Press, Chicago, 1987, pp. 25-40.

Simon, H. A., “The State of Economic Science,” in Sichel, W. (ed.), *The State of Economic Science. Views of Six Noble Laureates*, W. E. Upjohn Institute for Employment Research, Kalamazoo, MI, 1989, pp. 97-110.

Simon, H. A., “Prediction and Prescription in Systems Modeling,” *Operations Research*, v. 38, (1990), pp. 7-14. Reprinted in Simon, H. A., *Models of Bounded Rationality*. Vol. 3: *Empirically Grounded Economic Reason*, The MIT Press, Cambridge, MA, 1997, pp. 115-128.

Simon, H. A., “Introductory Comment,” in Egidi, M. and Marris, R. (eds.), *Economics, Bounded Rationality and the Cognitive Revolution*, E. Elgar, Aldershot, 1992, pp. 3-7.

Simon, H. A., “Colloquium with H. A. Simon,” in Egidi, M. and Marris, R. (eds.), *Economics, Bounded Rationality and the Cognitive Revolution*, E. Elgar, Aldershot, 1992, pp. 8-36.

Simon, H. A., “Altruism and Economics,” *American Economic Review*, v. 83, n. 2, (1993), pp. 156-161.

Simon, H. A., *The Sciences of the Artificial*, 3rd ed., The MIT Press, Cambridge, 1996.

Simon, H. A., “Bounded Rationality in Social Science: Today and Tomorrow,” *Mind and Society*, v. 1, n. 1, (2000), pp. 25-39.

Smith, A., *An Inquiry into the Nature and Causes of The Wealth of Nations*, W. Strahan and T. Cadell, London, 1776. Edited by Edwin Cannan with preface by George J. Stigler, The University of Chicago Press, Chicago, 1976.

SOLOW, R., *The Rate of Return and the Rate of Interest*, The Industrial Institute for Economic and Social Research, Stockholm, 1991.

TOULMIN, S., *Foresight and Understanding. An Enquiry into the Aims of Science*, Indiana University Press, Bloomington, 1961.

WINNIE, J. A., "Computable Chaos," *Philosophy of Science*, v. 59, n. 2, (1992), pp. 263-275.

II

Epistemological Issues Related to a General Framework

Problems of Determinism: Prediction, Propensity and Probability

Peter Clark

1. Introduction

There is perhaps no other metaphysical doctrine which has generated such a vast literature and heated controversy as the thesis of determinism. Determinism is a classic example of how a metaphysical problem is thrown up by the development of Science, in particular by classical, relativistic and quantum physics. Currently the concept is of much interest in two areas of research. Firstly in the foundations of physics where it is a question of real difficulty and depth as to whether and to exactly what degree the quantum theory is a genuinely indeterministic theory. The issue centrally concerns two "interpretations" of quantum theory which do not involve the second form of state evolution in quantum mechanics involving collapse of the wave packet at measurement, namely Bohmian mechanics and the many worlds interpretation of Everett and Dewitt.[1] While within the corpus of statistical physics there remains the long standing issue, introduced into physical theory in the mid-nineteenth century by James Clerk Maxwell and Ludwig Boltzmann, as to how statistics is to be combined in a consistent, conceptually transparent way with mechanics. Taken together these problems raise starkly the issue of how it is possible to simultaneously satisfy the apparently completely contradictory constraints of indeterminism on the microphysical level with determinism on the macrophysical level. Further the thesis of determinism is central to understanding how statistical mechanics works and how it can possibly be consistent to add probabilistic postulates to an underlying deterministic mechanics which that theory does. I shall discuss this issue, the so-called paradox of deterministic probabilities, at some length below.

It is worthwhile, however, from the start to make clear that I shall be concerned with physical determinism. I shall not be concerned in any systematic way with the various forms of the doctrine of Fatalism. None of the forms of this later doctrine are in any way, in my view, connected with physical determinism. It is no consequence of determinism for example that no matter what we do, what decisions we make and actions we execute, that the outcome of the events, what in fact takes place will be unaffected by our efforts. In some theological versions of fatalism appeal is made to the classical characterisation of God as having among other "maximal" properties, the property of omniscience. Since God knows everything, God knows the course of future events. But if God knows what future events will occur then what future events will occur is fixed and definite, for if it is the case that God knows that p, then (since knowledge is factive) p is true, for any proposition

[1] There is an excellent general introduction to the problem of the interpretation of quantum mechanics to be found in Putnam, H., "A Philosopher Looks at Quantum Mechanics (Again)," *The British Journal for the Philosophy of Science*, v. 56, (2005), pp. 615-634. See also Bohm, D., "A suggested Interpretation of the Quantum Theory in terms of 'Hidden Variables:' I and II," *Physical Review*, v. 85, (1952), pp. 166-193; Everett III, H., "Relative State Formulation of Quantum Mechanics," *Reviews of Modern Physics*, v. 29, (1957), pp. 454-462, and Dewitt, B., "Quantum Mechanics and Reality," *Physics Today*, v. 23, (1970), pp. 30-35.

p describing events. So for such a statement *p* describing some aspect of an event in the future, then it looks as if what is described by *p* is fixed and determined now. No doubt there is much that can be said about this form of argument. For our purposes, however, it is enough to make two remarks. First there is something of an equivocation over the concept of "knowing that" and secondly that even if the Universe were indeterministic the argument, if successful at all in the deterministic case, would still apply. The equivocation on "knowing that" arises from the fact that whatever "knowing that" consists in for omniscient beings, it is certainly a very different sense from "knowing that" for beings like us, or coming to know via the methods of science. For beings like us, who are not omniscient, knowing that *p* involves forming a belief on the basis of some form of evidence and justification. Concerning the course of events, we acquire beliefs on the basis of a temporal process, and justification of claims about the course of future events is given in general on the basis of the inductive discovery of regularities in nature and beliefs about the way things are at a particular time. Thus belief acquisition is a temporally directed process. But the sense of God's supposed omniscience is decidedly atemporal; it is, to employ a metaphor, without or "outside" time. Given this supposition in no sense is God's knowledge of events in our future, *fore*knowledge of those events. There is no "now" or "present" for that knowledge to ante-date.

Suppose we take as a model of a simple indeterministic Universe the successive throws of a fair coin. Let us consider an infinite sequence of throws such that for each throw the outcome at any throw is independent of that of the previous throw. The result of this is a countably infinite sequence of zeros and ones, which can serve as a model of the history of "events" (i.e. successive outcome zero or one) in our toy indeterministic Universe. In so far as the notion of an all-knowing being can be given any sense, it is clear that such a being could know each and every outcome of this infinite sequence in the atemporal sense without it being the case that an "observer" whose knowledge was confined merely to some initial segment of the sequence could discern any projectable pattern there.

It is far from clear that the theological notion of an omniscient knower in the above sense is itself a coherent conception. A more reasonable notion of omniscience might be that of a knower who no matter how much we know, knows more, or perhaps of the knowledge embodied at the limit of scientific activity, when all scientific questions have received an answer. But on neither of these conceptions is there any, even apparent entailment of physical determinism, for neither of these conceptions requires that the omniscient knower knows *all* the events in the Universe's history.

Neither does the quasi-logical form of fatalism follow from, nor entail, determinism. The claim that it does entail determinism for the class of events in the history of the Universe has two forms. One might be dubbed the "whatever is, always was to be" argument and the other the "bivalence" argument from future contingents. The idea behind "whatever is, always was to be" argument is this. Consider the statement 'Andrew Wiles proves Fermat's Last theorem in September 1994.' This statement is true. Then, so we are invited to reason, it must always have been true even a million years ago that in that certain year, Wiles would prove the theorem. So "whatever is", *viz*. in our example Wiles proving what had formerly been merely a conjecture, "always was to be", since the statement 'Andrew Wiles proves Fermat's Last Theorem in September 1994' if true at all, must have been true at any time prior to the event described by the statement (and for that matter any time subsequent to it). Is not this a kind of determinism: if it was true a million years ago that Wiles would prove the famous conjecture, was not he determined so to do?

The answer is rather clearly not, for we can see that all the premises of the argument can be true in an indeterministic universe. All that they demand is the truth of the claim that Wiles did prove Fermat's last theorem in 1994 and a claim about the timelessness of truth or falsity as applied to statements. This latter is essentially a semantic thesis, a thesis about what we mean when we claim that a statement is true and holds quite independently of whether our Universe is deterministic or indeterministic. If we return to our toy indeterministic model universe above, we can see that for each index in our infinite sequence of zero's and one's, only one of these values occurs. Let us imagine that zero occurs at say the k th index position. So it is true that as a matter of fact "zero occurs at index k" and if at throw i, for i strictly less than k an observer had uttered the statement (for whatever reason —a bet perhaps) "zero occurs at throw k" they would have uttered, in virtue of the thesis of the timelessness of truth, a true statement. But that is completely compatible with the claim that the events at i fail to fix or determine the events at throw k, as indeed we have supposed to be the case.

A similar point undermines the argument from future contingencies to determinism. This venerable argument says that since every statement is either true or false but not both,[2] if p is a statement describing some future event either p is true or not p is true. We may not know which it is, but one of them is. So it is now determined, of whichever one it is (say p) that p is true, so whatever p says will occur. But again this argument hinges on mixing semantic facts, how we ascribe truth values to sentences with physical facts.[3] And of course, though our toy model indeterministic universe satisfies bivalence, *i.e.* each indexed throw has a "zero" or a "one" and not both as the outcome at that throw (by supposition we exclude the coin landing on its edge) the throws are independent, so this tells us nothing about whether or not any initial segment of the sequence determines or fixes any or all subsequent members of it.

A second version of determinism which is sometimes connected in the literature with physical determinism is the notion of causation. The connection between the two is supplied by some principle like "every event has a cause and like causes produce like effects" but if we actually attend to the content of physical theory the notion of causation plays little role if any in the foundations of physics. Much more seriously, however, it does not seem to be in any way correct to identify *global causality* with physical determinism. The reasons for this are two-fold. First, there are circumstances which one might very well want to regard as indeterministic (intuitively, at least), which are nevertheless ones

2 This venerable argument goes back at least as far as ARISTOTLE, *De Interpretatione*, Chapter 9. English Edition by J. Barnes in *The Complete Works of Aristotle*, The Revised Oxford Translation, Volume 1, Princeton University Press, Princeton, NJ, 1984, pp. 28-30.

3 This view is sometimes characterised as the "thin red line" account of the compatibility of objective indeterminism with the determinate truth values of future contingents, see particularly LEWIS, D., *On the Plurality of Worlds*, Basil Blackwell, Oxford, 1986, pp. 206-209. For a very different view which argues that this leaves no space for a genuine sense of possibility and thus is just determinism after all see BELNAP, N., *Facing the Future*, Oxford University Press, Oxford, 2001, Chapter 6, and MACFARLANE, J., "Future Contingents and Relative Truth," *Philosophical Quarterly*, v. 53, (2003), pp. 321-336.

in which causal factors are present the well known example of lung cancer and cigarette smoking being a classic case. Secondly, there are situations which according to physical theory are deterministic but are acausal.

A simple case is provided by Newton's First Law. According to it, a particle will remain in a state of rest or of uniform motion in a straight line unless it is acted upon by an external impressed force. This inertial motion of a single particle moving under no forces in an otherwise empty universe, is a paradigm case of deterministic motion, if anything is. Yet it would be wrong to regard this motion as caused or causal. All physical theories of motion begin with an account of natural motion —motion which, requires no explanation at all, and certainly requires no causal explanation. In Aristotelian physics, natural motion is the motion of a body toward its natural place.[4] Aristotelian physics sought to render intelligible unnatural or violent motion, that is motion which consists of displacement from the natural place, that does require a cause. In Newtonian physics, on the other hand, natural motion is inertial motion along a geodesic in space-time, which of course in the Newtonian case is a Euclidean straight line. In relativity theories, the natural motion of a body is again along the geodesic paths through space-time. In all of these cases, we have a situation in which a certain motion is postulated as natural and all other motion is explained by recourse to causal factors which account for the deviation from the natural state. So the situation in the case of natural motion is that we have completely deterministic motion, yet motion which is acausal. Causes are introduced if at all by the theory only to explain deviation from the natural state of motion, while nonetheless, the natural motion is certainly deterministic.

2. Laplacean Determinism

It is traditional in the study of physical determinism to begin with the famous conception of determinism due to Pierre Simon, Marquis de Laplace. Actually he deploys two quite separate notions of determinism, first he uses the Principle of Sufficient Reason and second he identifies determinism with a series of very strong claims about predictability. In the introduction to his short treatise on the philosophical foundations of probability theory in which he introduced the classical interpretation of the probability calculus he argued as follows:

"Present events, are connected with preceding ones by a tie based upon the evident principle that a thing cannot occur without a cause which produces it. This axiom, known by the name of the *principle of sufficient reason*, extends even to actions which are considered indifferent; the freest will is unable without a determinative motive to give them birth; if we assume two positions with exactly similar circumstances and find that the will is active in the one and inactive in the other, we say that its choice is an effect without a cause. It is then, says Leibnitz, the blind chance of the Epicureans. The contrary opinion is an illusion of the mind, which, losing sight of the evasive reasons of the choice of the will in indifferent things, believes that choice is determined of itself and without motives.

We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for instant an intelligence which

---

[4] Cf. Aristotle, *Physics*, Book Five. English Edition by J. Barnes in *The Complete Works of Aristotle*, The Revised Oxford Translation, Volume 1, Princeton University Press, Princeton, NJ, 1984, pp. 378-390.

could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it —an intelligence sufficiently vast to submit these data to analysis— it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future as the past, would be present to its eyes. The human mind offers, in the perfection it has been able to give to astronomy, a feeble idea of this intelligence. Its discoveries in mechanics and geometry, added to that of universal gravity, have enabled it to comprehend in the same analytical expressions the past and future states of the system of the world. Applying the same method to some other objects of its knowledge, it has succeeded in referring to general laws observed phenomena and in foreseeing those which given circumstances ought to produce. All these efforts in the search for truth tend to lead it back continually to the vast intelligence which we have just mentioned, but from which it will always remain infinitely removed."[5]

The constraints that Laplace imposes on determinism as he formulates it in the second paragraph of the quotation are very strong. There are four of them put in modern terms: first that an *analytic* solution to the equations of motion for a system obeying Newtonian mechanics *expressible as a closed formula exists* for *all* initial conditions (i.e., the equations of motion will be integrable, as Laplace has it, his demon would provide solutions which "would embrace in the *same formula* the movements of the greatest bodies of the Universe and those of the lightest atom"). The second constraint is that the solutions to the equations of motion will be *global*, that is will generate solutions for all time, not just *local* solutions. The third constraint amounts of the requirement that every solution to the equations of motion for such a system shall be *effectively computable* in the data, that is, given as initial conditions the position and momentum (the values of the state variables) of all the particles in the Universe at some given instant, the values of the state variables at any subsequent instant shall be effectively computable functions of the state variables at the arbitrarily chosen initial instant (i.e. as Laplace has it that the solution should enable one to "foresee those [phenomena] which given circumstances ought to produce"). The fourth constraint is one of the complete accessibility of the data of initial conditions, that is, it is to be assumed in the formulation of the deterministic claim that it is always possible (classically at least) for experimental evidence to fix an exact real number as the value of each state variable at an instant so that there is no uncertainty in the initial data. This means that no error in the initial data exists which may increase as time goes on eventually destroying the possibility of accurate prediction at all.

No doubt if these very strong conditions were to obtain then Laplacean global predictability would obtain,[6] but there is no reason at all to tie the claim of physical determinism to a thesis of global predictability. Indeed in general in classical mechanics

[5] Pierre Simon, Marquis de Laplace, *A Philosophical Essay on Probabilities*, 1820. translated from the Sixth French edition by F. W. Truscott and F. L. Emory, Dover Publications, New York, 1952, pp. 3-4.

[6] These Laplacian constraints are certainly not inconsistent. An interesting example of a model of them is provided by the only known fully general solution to n-body problem, that is for a system of N material particles moving under an attractive force law acting along the line of centres but proportional to the distance of separation. In this special case there is a frame of reference in which the equations of motion are separable and are solvable individually. See Pollard, H., *Mathematical Introduction to Celestial Mechanics*, Carus Mathematical Monographs, v. 18, 1976, p. 59, Exercise 1.3.

these Laplacean conditions cannot be satisfied. Very many equations of motion especially non-linear ones are known not to be integrable, no solution expressible in a closed form exists for them. Such systems are at heart of chaos theory. Similarly local solutions to the equations of motion may exist for short time intervals but these may not be extendable to global solutions because of the existence of singularities. An easy example is provided by a point mass moving along a straight line subject to a damping force which varies as a power of the velocity. Thus when the force is proportional to v^k for $k < 1$, there are no unique solutions for all times. Further there are now known to be examples of non-computablity in classical physics.[7] But perhaps the central reason why it is quite wrong to identify predictability with determinism is the existence of deterministic chaos. In Nineteenth century physics the dominant paradigm was that found in perturbation theory especially in celestial mechanics, which essentially asserted that small perturbations in a motion produced small effects on that motion, that is to say that if one had a small error in initial conditions that error would propagate only linearly with time. So even with unstable motions like that produced by the gravitational interaction of many asteroids or comets with the orbit of a planet the very small error produced by neglecting the interactions on the predicted planetary orbit would be correspondingly small and would as such yield significant error only after a very long time. However with chaotic systems which are in every natural sense deterministic, the situation is very different. For those systems uncertainty in the initial values propagates exponentially with time. Such extreme sensitivity to initial conditions means that after a very short time the exact state of the system may become very uncertain indeed if the initial condition is not known with complete accuracy.[8] One can say in conclusion then that predictability and determinism have more or less completely come apart.

3. Deterministic Theories

Let us ask the question, "what do physicists really mean when they describe a physical system as deterministic?" It is that the state at one time fixes the state at future times. What this means for mechanics is the well known fact that because of the existence and uniqueness of solution of the differential equations of motion the time evolution of a dynamical system (say a system of moving molecules, the familiar billiard balls of a dilute gas) is represented by a unique trajectory in phase space (the space in which a point corresponds to the instantaneous state of the system).

Through each point of the phase space there passes *at most* one trajectory, as a consequence of uniqueness of solution. In any time interval ∂t the point on the trajectory corresponding to the state of the system at t is transformed into the *unique* point on the trajectory corresponding to the state at $t + \partial t$. But of course during the same time interval *every* other point in the phase space is transformed uniquely by the solution to the differential equations of motion into another point in the phase space. Hence the solutions

---

7 See particularly Pour-El, M. and Richards, J., "Non-computability in Analysis and Physics," *Advances in Mathematics*, v. 48, (1983), pp. 44-74 also Pour-El, M. and Richards, J., *Computability in Analysis and Physics*, Springer-Verlag, Berlin, 1988.

8 An excellent technical introduction to chaos theory is Wiggins, S., *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, Springer-Verlag, Berlin, 1990. A good non-technical introduction can be found in Gleick, J., *Chaos: Making a New Science*, Viking Press, New York, 1987. Of course not all aspects of the motion are unpredictable, but certainly the evolution of exact trajectories certainly are.

to the equations of motion define a map from the phase space of the system into itself with the property that as the dynamical system evolves through time, that evolution induces a "*flow*" of the points of phase-space. The flow has the very important property that it characterises a one parameter (the time) semi-group, which means that if we consider any point in phase space and look at where the flow takes the point in time t and then subsequently in time interval ∂t, the final position of the point is exactly that to which the flow takes the point in time interval $t + \partial t$. This is the familiar *natural* motion of the phase space. (Since the equations of motion of classical mechanics are time-symmetric, the natural motion of the phase space has the structure of a one parameter group.)

As was noted by Montague in his classic paper of 1974 on deterministic theories this idea can be generalised in a natural way.[9] If we consider an *isolated* (this restriction is absolutely essential) physical system and think of the history of that system as the "graph" of the values of its state variables in phase space (i.e. the n-dimensional space in which the instantaneous state of the system is a "point") through time, then the system is *deterministic* if and only if there is one and only one possible path consistent with the values of its state variables at any arbitrary time. There are however four basic problems in formulating this intuitive condition precisely. One concerns the appeal to physical *systems*, the second is the notion of *state* and the third is the restriction to *isolated* systems and the fourth is the reference to *possible* path.

The first difficulty stems from the fact that systems can exhibit deterministic and indeterministic aspects simultaneously. An obvious example is a system obeying classical mechanics if we consider only the temporal evolution of its observational states. Since an observational state will in general have a dynamical image (i.e. the set of phase points compatible with the observational state) containing more than one phase point, two systems obeying the same dynamical laws in the same observational state at any given time may well be found in different observational states at later times. It is precisely this characteristic which gives the deterministic thesis its *hidden variable* aspect; for a claim that a system is indeterministic may rest merely upon our not having obtained a description of the underlying state-variables, i.e. the theory of the system may be radically *incomplete*. This illustrates just how difficult it may be to ascertain whether a particular system is deterministic or not. For the claim that a system is indeterministic is always open to challenge that our description of its state is simply incomplete.

A second difficulty lies in the appeal to the notion of the *state* of the system itself. This too is a highly theory dependent notion. In classical mechanics in its simplest form the state of a moving particle is characterised by the pair <position, momemtum>, in thermodynamics the state of an ideal gas is specified by say the pair <temperature, volume> and in classical electromagnetism the state of the field (in empty space) is given by value at any point of the pair <electric field vector, magnetic field vector>. In each theory differential equations (expressing the fundamental laws of the theory) lay down

9 Montague, R., "Deterministic Theories," in Thomson, R. H., *Formal Philosophy*, Yale University Press, New Haven, 1974, pp. 302-359. The *locus classicus* for all treatments of the notion of physical determinism is Earman, J., *A Primer on Determinism*, Riedel, Dordrecht, 1986. See also Fraassen, B. van, "A Formal Approach to the Philosophy of Science," in Colodny, R. G, (ed.), *Paradigms and Paradoxes*, The University of Pittsburgh Press, Pittsburgh, 1972, pp. 303-366.

how such states evolve with time. But the notion of state must not be sufficiently general to allow determinism to be trivially true, they must not for example code information about how the system will develop in the future.[10] Clearly the notion of state must be closely connected with that of an intrinsic property but it is very difficult to say exactly what makes a given property intrinsic.

The third difficulty concerns the restriction to isolated systems. Clearly this restriction is absolutely necessary since no system can satisfy the uniqueness of solution condition if it is subject to interference from outside. We might well have a unique solution to a planetary orbit for a considerable period of time at least but not if it is hit by a planetary asteroid from without the solar system. The trouble is of course that no system can be completely isolated from outside effects. To compensate for this we might very well consider an expanded system made up of our original system and its environment. But again the expanded system will not be isolated and we will need to treat of a still larger system composed of the expanded system and now its environment. Soon we will have encompassed the entire Universe. So that is how we will treat of a physical system and always think of it as isolated in its own "Universe" containing nothing else, but the system in question.

The fourth difficulty concerns the introduction of the modality, possibility, and the attendant panoply of real but not actual possible words. One way of avoiding this is to think of the class of all models of the theory. To cut a quite long story short, *a theory T* is said to be *deterministic in the state variables* (say $\partial_1, \ldots, \partial_n$) if any two standard models, or *histories*, as they are called, of the theory (i.e. any two relational structures which are models of the theory) which agree at some given time, agree at all other times. In short the constraint entails that for a deterministic theory if two histories (or models) have identical states at one time then they have identical states at all times[11]. A physical system may be said to be deterministic in the state variables ($\partial_1, \ldots, \partial_n$) when its history *realises* (or satisfies) a *theory* deterministic in the state variables ($\partial_1, \ldots, \partial_n$). This characterisation of a deterministic theory and deterministic system fits very well with classical, relativistic and quantum mechanics. It has the further advantage of not relying on the essentially extraneous notion of predictability at all. As is now well known given this characterisation of determinism, Newtonian mechanics is not in general deterministic,[12] while as we noted above there are versions of the quantum theory that most certainly are.

[10] Jeremy Butterfield has given a nice example of what has to be excluded in the notion of state if determinism is not to be trivially true. He writes: "Thus, to take an every day example, 'Fred was mortally wounded at noon', implies that Fred later dies. But the property ascribed at noon is clearly extrinsic: it 'looks ahead' to the future. And so this implication does not show that there is any genuine determinism about the processes that led to Fred's later death." Butterfield, J., "Determinism and Indeterminism," *Routledge Encyclopedia of Philosophy*, v. 3, (1998), pp. 33-38.

[11] We think of a history of a physical system as roughly the course of values that its states take in a universe which contains only that system, as it evolves through time. The system is deterministic if any two such histories agree at one time then they agree at all times. We assume that the basic laws are time symmetric. It might be objected that we have assumed some sort of meta-time across these universes so that we can talk about states agreeing at a time. But actually we can do without such talk. We can replace it by saying that if at an instantaneous time slice of the history of system A coincides with an instantaneous time slice of system B, then the subsequent (and past) time slices in each of the systems will be coincident in the values of their state variables.

[12] See particularly Earman, J., *A Primer on Determinism*, pp. 33-40, and Xia, Z., "The Existence of Non-collision Singularities in Newtonian Systems," *Annals of Mathematics*, v. 135, (1992), pp. 411-468.

4. Statistical Mechanics and the Paradox of Deterministic Probabilities

One of the most fascinating problems for the philosopher posed by the corpus of statistical physics is the issue of the consistency problem which arises in both the classical and the quantum contexts. It arises starkly in classical statistical physics, there it is the issue of precisely how it is possible to add probabilistic assumptions interpreted as physical probabilities, describing chances in Nature to treat of an aggregate motion, the component sub-motions of which, being governed by the laws of Hamiltonian mechanics, are entirely deterministic.[13] Essentially the problem occurs because one of the two theories we want to employ viz. Hamiltonian mechanics is a completely general theory, that is it ought to give a complete description of any physical situation to which it applies, hence if we put them together the suspicion must be that they will over-determine the history of the physical system under consideration and inconsistency will result.[14]

To explain the macroscopic phenomena associated with thermodynamics we need to introduce stochastic postulates "on top of" Hamiltonian mechanics, but the latter mechanics gives a complete state description of the physical situation and all the values of the observables are thereby determined as functions of the variables describing the instantaneous state. The situation is of course compounded in the case of non-equilibrium statistical mechanics. The reason is that in theories of this kind we want to use deterministic and time-symmetric theories (once again classical dynamics) to obtain an explanation of the well-known phenomenon that the behaviour of physical systems that are far from equilibrium is not symmetric under time reversal. That is we want to obtain explanations of facts the type subsumed by the second law of thermodynamics. Since there is no possibility whatever of combing time-symmetric deterministic dynamics to obtain, even in the limit, non-time-symmetric theories; some postulate of a stochastic kind (e.g. that the hypothesis of molecular chaos must hold for some part of the time in the approach to equilibrium, or some postulate asserting the very special nature of the initial conditions) must replace the time-symmetrical dynamics.

As is very well known despite the enormous empirical success of statistical mechanics the conceptual problems surrounding it have persisted.[15] They are particularly acute if one wishes to claim that the stochastic assumptions deployed in statistical mechanics describe physical facts, just as much as does the underlying dynamics. That is to say the stochastic

---

[13] This is because the equations of motion for the dynamical system under consideration, say a dilute gas (the canonical Hamiltonian equations, where the Hamiltonian is independent of the time) are such that we have a system of first order differential equations whose solutions *always exist* and are *unique*, given an arbitrary initial point in phase space at some instant *t*. Since both existence and uniqueness of solution of the differential equations of motion obtain, the dynamical states of the system are determined.

[14] The history of the kinetic theory of gases provides examples where the suspicion was more than well-founded, two of the most notorious examples are Maxwell's original assumption in his 1860 paper that the components of velocity of a single moving molecule are statistically independent thereby contradicting the law of the conservation of energy and Boltzmann's original claim that what is now called the hypothesis of molecular chaos holds all of the time in his first derivation of the H-theorem, contradicting the time-symmetry of classical mechanics. Cf. Maxwell, J. C., "Illustrations of the Dynamical theory of gases," *The Scientific Papers of James Clerk Maxwell*, edited by W. D. Niven, Dover Publications, New York, 1965, pp. 377-409 and Boltzmann, L., "Further Studies on the Thermal Equilibrium of Gas Molecules," in Brush, S. G. (ed.), *Kinetic Theory, v. 2, Irreversible Processes*, Pergamon Press, Oxford, 1966, pp. 88-175.

[15] For a superb overview of the central conceptual problems of statistical mechanics see Sklar, L., *Physics and Chance*, Cambridge University Press, Cambridge, 1993.

postulates of the theory describe objective physical properties of the systems under consideration just as much as the underlying dynamics does, no more and no less. There is certainly a general view that this is just impossible and that if one has an underlying deterministic dynamics then the probabilistic statements added to it must be interpreted merely as subjective statements expressing only degrees of belief. This I think is wrong, but the argument for it looks very compelling. Loewer has put the point very nicely:

> "If the laws are deterministic then the initial conditions of the Universe together with the laws entail all facts —at least all facts expressible in the vocabulary of the theory. But if that is so there is no further fact for a probability statement to be about. Someone who knew the laws and the complete initial condition —a Laplacean demon— would be in a position to know everything. So what else can the probabilities posited by such theories be but measures of our ignorance. And since it is assumed we know the laws that ignorance must be ignorance of initial conditions."[16]

So the argument amounts to this: if the system is deterministic then, if we really knew the values of the state variables exactly, every probability would in reality reduce to zero or one, so all non trivial probabilities would merely arise from ignorance of exact states. If $<q_1(t), \ldots q_n(t)\, p_1(t), \ldots p_n(t)>$ is the instantaneous dynamical state at some time t of one of the types of system studied in statistical mechanics and X is some integral function of the instantaneous state variables, the value of X at any time being an observable property of the system, then the conditional probability that X has the value, say k, at some time $T > t$ is

$$P(X(T) = k \,/ < q_1(t), \ldots q_n(t)\, p_1(t), \ldots p_n(t) >) = 1 \text{ or } 0.$$

So if we conditionalise on the exact state in a deterministic system we will always get either one or zero as a result. This claim is true, but does it mean that every probability statement for such theories which is non-trivial must be measuring merely our ignorance as to exact states. Some distinguished philosophers like Popper thought it did. He wrote:

"Today I can see why so many determinists, and even ex-determinists who believe in the deterministic character of classical physics, seriously believe in a subjectivist interpretation of probability: it is in a way the only reasonable possibility which they can accept: for objective physical probabilities are incompatible with determinism; and if classical physics is deterministic, it must be incompatible with an objective interpretation of classical statistical mechanics."[17]

Similarly David Miller remarks in the context of defending a propensity interpretation of probability that "for there to exist probabilities other than zero or one it is therefore essential that the class of future possibilities contains more than a single element,"[18] which of course cannot be so in a deterministic universe.

I will consider three ways one might seek to give an objective interpretation of probability in statistical mechanics; a propensity theory consistent with the underlying

[16] Loewer, B., "Determinism and Chance," *Studies in History and Philosophy of Modern Physics*, v. 32, n. 4, (2001), pp. 609-620. Like me he does not accept it, though he criticises the general approach that I used in an earlier paper in an attempt to diffuse the argument. Clark. P., "Determinism and Probability in Physics," *Proceedings of The Aristotelian Society*, Suppl. Vol. 62, (1987), pp. 185-210.

[17] Popper, K. R., *Quantum theory and the Schism in Physics*, Routledge, London, 1992, p. 105.

[18] Miller, D., *Critical Rationalism: A Restatement and Defence*, Open Court, La Salle, IL, 1994, p. 186.

dynamical determinism, the relative frequency view and ergodic theories and a view related to that of David Lewis recently proposed by Barry Loewer.

The way in which Popper and Miller think about propensities is just one way to conceive of them and it is certainly possible to think of propensities, as Popper sometimes wrote as if he did, as simply tendencies or dispositions in the mechanisms which produce the sequence of trials with chance-like outcomes to generate those given outcomes. In this latter weaker sense of propensity, it certainly doesn't follow that there is any immediate conflict with determinism.

This is clearly seen in recent work by Patrick Suppes. He proves two most interesting representation theorems, both for systems with underlying deterministic mechanics. The first example is a model of coin-tossing where a coin of negligible thickness has initial vertical velocity u and angular velocity ω. The result he establishes is that "essentially any mathematically acceptable probability distribution of u and ω, at time $t = 0$, will lead to the probability of heads being approximately 0.5. The mechanical process of tossing dominates the outcome. Small variations in u and ω which are completely unavoidable, lead to the standard chance outcome. Put explicitly, the standard mechanical process of tossing a coin has a strong propensity to produce a head as outcome with probability 0.5"[19] Of course here we begin with a probability distribution over the initial conditions but what the analysis shows is that the system has a tendency whatever that initial distribution is to almost always generate heads with probability 0.5 given the dynamics of the system. The second theorem which Suppes uses is an example of the restricted three body problem in mechanics which involves no collisions whatever and is thus entirely deterministic but which nevertheless exhibits in a parameter of its motion a fully random sequence, for certain sets of initial conditions. He rightly says of the example that: "In the ordinary sense, as reflected in the theory of differential equations, in the philosophical talk of determinism, etc., the three body problem described is purely deterministic in character and yet randomness of any desired sort can be generated by appropriate initial conditions. The wedge that has sometimes been driven between propensity and determinism is illusory."[20]

Now it is important to note that Suppes is not proposing a general propensity interpretation of probability but rather pointing out in an illuminating way how the generating mechanisms in certain specific cases produce the chance like character they do and the way they do it.

We are engaged in trying to understand how physical probability, which is just as objective as mechanics, can be understood in deterministic contexts. For that reason we will not be considering subjective theories of probability which characterise probability as a measure of degree of belief because it is very difficult to see how statements about degrees of belief can play a law-like role in explaining phenomena like those explained in statistical mechanics. What role can degrees of belief play in explaining why kettles boil when heated or baths settle down to a pleasant uniform temperature when over-filled with very hot water or why my glass of whisky and ice settles down to a refreshing drink rather

---

[19] Suppes, P., *Representation and Invariance of Scientific Structures*, CSLI Publications, Stanford, 2002, pp. 21-225. The quotation is from p. 216.

[20] Suppes, P., *Ibid*., p. 222

than evolve into an unpleasant mixture of colder ice cubes and hotter whisky? These are law like behaviours exhibited in Nature, it is difficult to see how these regularities in Nature can be explained by reference to my or anyone's degree of belief. However if we are to give an interpretation of physical probability it had better be an interpretation of probability, that is had better satisfy the axioms and thereby theorems of the probability calculus and there is good reason to believe that propensity theories in general to not do that. As is well known the problem, first pointed out by Paul Humphreys and Wesley Salmon,[21] concerns the existence of inverse probabilities which always exist when the conditional probability is defined, but can the inverse propensity always exist? Not in general, especially not if propensity is thought of as some form of partial causation or generalised force. In a recent article Humphreys after surveying a number of candidate general accounts of propensity has concluded that they cannot be a satisfactory interpretation of probability theory.[22] Interestingly he concludes his article by raising the possibility that "probability theory does not have the status of a universal theory of chance phenomena with which many have endowed it."[23] This is a view which I shall not pursue here since we are concerned with standard probability theory as it is used in statistical physics.

The second interpretation we should consider is that of the frequency theory. On this theory the probability of an event of type B in an infinite sequence trials generated by a particular process of type A is the limit of the relative frequency of B's occurrence in the sequence as the number of trials tend to infinity. But two conditions must be satisfied by the sequence. First the limit of the relative sequence must exist and the sequence must be random with respect to the appearance of B. What is of particular interest here is how the randomness requirement is to be understood, since that will constrain whether deterministic sequences can satisfy the second condition and allow probabilities to be defined. But when is a sequence of outcomes random?

One famous solution to this problem due to Alonzo Church and Richard von Mises makes very precise the idea of a sequence being random.[24] Consider the infinite sequence of outcomes $\{A_o, A_1, A_2, A_3, \ldots , A_n, \ldots\}$ such that some of the A_i have a property, say B. We want to say under what circumstances B occurs randomly in the sequence, or when the sequence is random with respect to the property B. First ask, what would it be for B to occur non-randomly? This would mean that we could find a rule which allowed us to *anticipate* the occurrences of B in the sequence —perhaps that, B occurs, in the outcome, every 10th trial, or B occurs every time the number of the trial is even, and so on. In other words, occurrences of B in the sequence $\{A_n\}$ would be governed by a rule, and if we selected a subsequence simply by using this rule and selecting just those members of the main sequence which the rule told us would have the property B, we could thereby obtain a sequence in which only outcomes with property B occurred. So the relative frequency of

---

21 SALMON, W. C., "Propensities a Discussion Review," *Erkenntnis*, v. 14, (1979), pp. 183-216; and HUMPHREYS, P., "Why Propensities cannot be Probabilities," *The Philosophical Review*, v. 94, (1985), pp. 557-570.

22 HUMPHREYS, P., "Some Considerations on Conditional Chances," *The British Journal for the Philosophy of Science*, v. 55, (2004), pp. 667-680.

23 HUMPHREYS, P., "Some Considerations on Conditional Chances," p. 679

24 Cf. CHURCH, A. "On the Concept of a Random Sequence," *Bulletin of the American Mathematical Society*, v. 46, (1940), pp. 130-135, and VON MISES, R., *Probability, Statistics and Truth*, Dover Publications, New York, 1957.

B in this newly selected subsequence would be one, even though the relative frequency of B in the main sequence might be, for example, one half.

According to the Church-von Mises conception the randomness of the sequence should rule such a possibility out. A random sequence would then be one in which every rule for selecting sub-sequences from the main sequence has the property that the relative frequency of the characteristc B in the selected subsequence is the same as the relative frequency of B in the main sequence. In short there should be no way of anticipating the occurrence of B in the sequence, for if there were we could use it to select a subsequence in which the relative frequency of B's would be 1, no matter what it was in the main sequence.

However, the notion of "rule" employed here is not satisfactory, since it has not yet been made clear what sort of procedure we would regard as an acceptable rule —what should be allowed as a rule and what should not? Church provided a way of answering this question. A rule should be any algorithm or effectively computable procedure for selecting the subsequence from the main sequence. So, a rule means an effectively computable function whose argument place is filled by the number corresponding to the position of the member of the main sequence and whose value is a yes/no decision as to whether to include or exclude that member of the main sequence from the selected subsequence. Such a procedure systematically generates a subsequence of the main sequence, whose members are exactly those for which the function gave the answer "Yes."

A random sequence, therefore, in the Church-von Mises sense is one for which there is no effectively computable function which selects a subsequence in which the frequency of the characteristic in question is different from the relative frequency of that characteristic in the main sequence. Unfortunately this account is too liberal and too strict. It is too strict since it rules out any finite sequence from being random, we could simply effectively list any desired subsequence. It is also too liberal in that it allows some sequences to be random which we might certainly want to discount. An infinite sequence of zero's and one's with an initial segment of a million ones could still turn out to be random on this view. Worse, by a theorem of Ville,[25] there are sequences of zero's and one's with the property that though in any sub-sequence selected by a computable rule the limit of the relative frequency of 1 is one half, in *every* finite initial segment of the sequence the frequency of ones is greater than a half. A very important extension of Von Mises idea was provided by Martin-Löf.[26] He showed that the non-random sequences of zero's and one's form a set of measure zero in the set of all binary sequences. So intuitively a selection from the set of all binary sequences would yield a random sequence with probability one.

An approach to the characterisation of randomness which avoids some of the difficulties but which retains the connection between non-randomness and algorithmic procedures for anticipating the outcome of trials is that of Kolmogorov and Chaitin. This idea hinges on the notion of the complexity of a sequence. Suppose we have a finite sequence of zeros

[25] Cf. VILLE, J., *Etude Critique de La Notion Collectif*, Gauthier-Villars, Paris, 1939.

[26] Cf. MARTIN-LÖF, P., "The Definition of a Random Sequence," *Information and Control*, v. 9, (1966), pp. 602-619.

and ones (corresponding to throwing a fair coin a thousand times, writing zero if a head appears and one if a tail does). One possible such sequence is the following.

<0, 1, 0, 1, 0, 1, ... 0, 1, ... , 0, 1>

If we were trying to write a computer program to print out this binary sequence, we would need only a very "short" program, for the sequence is generated by a very simple rule; put zero in all odd places and 1 in all even places in the sequence. If, however, we think of the sort of sequence which would actually be yielded by tossing a coin, we would need a program as long as the sequence, in order to have the computer print out the sequence. Intuitively the sequence would be so complex, so disordered that the only way by which our computer could generate the sequence would be by being supplied with either the sequence itself or some set of instructions as long and complex as the sequence. If we think of a computer as a particular way of representing an algorithm, we can define the complexity of a sequence with respect to a given algorithm as the shortest number of steps needed to generate the sequence using the algorithm. This account of course makes the complexity of a sequence depend upon the algorithm chosen, but an important theorem of Kolmogorov can be employed to remove this dependency on the particular algorithm chosen. When a sequence has a complexity of the same order as the length of the sequence, it is said to be maximally complex and it is Kolmogorov's claim that the maximally complex sequences in the sense specified above are exactly those sequences which are random in the intuitive or informal sense. On this view an infinite sequence is random just when all its initial segments are maximally complex finite sequences.[27]

One thing is clear from both the above accounts of the nature of a random sequence and that is that nothing in the notion of a sequence of outcomes as being random ensures that the sequence of outcomes cannot be deterministic in the Montague-Earman sense. It is well-known that there are sequences which are Von Mises random which model a deterministic theory[28] and nothing in the characterisation of randomness as concerned with measure or complexity rules out that such sequences may be deterministically generated. So randomness is most certainly not indeterminism in disguise. This is very important because it allows for an intimate connection between the evolution of certain dynamical systems and the applicability of the frequency theory of probability. The connection is best seen in the case of "ergodic systems." For such systems a famous theorem of Birkhoff informs us of a very interesting statistical property: except for a set of initial points of trajectories in phase space of the system of measure zero, given any integrable function f of the state variables, consequently including those whose values determine the observational states of the system, the average value of f along the dynamical trajectory of the system

[27] An excellent account of this approach to the notion of randomness can be found in Chaitin, G. J., *Algorithmic Information Theory*, Cambridge Tracts in Theoretical Computer Science, Cambridge University Press, 1987. There Chaitin proves that for the real numbers the two basic definitions of random real (that of Martin-Lof and Kolomogorov) are equivalent. In an interesting article which contains a thorough survey of the literature on randomness Antony Eagle argues for an epistemic account of the notion based upon maximal unpredictability. While this might be an interesting notion it will hardly satisfy our purpose of providing an objective interpretation of randomness in statistical mechanics. It would once again make the explanatory nature of the probabilistic postulates of statistical physics depend upon our states of knowledge. See Eagle, A., "Randomness is Unpredictability," *The British Journal for the Philosophy of Science*, v. 56, (2005), pp. 749-790.

[28] Humphreys, P. W., "Is 'Physical Randomness' Just Indeterminism in Disguise?," *PSA* 1978, v. 2, pp. 98-113.

is equal in the limit as t tends to infinity to the average value of f determined over the hypersurface of constant energy in the phase space. This means that if the system is ergodic, as t tends to infinity its representative point will pass arbitrarily close to every point on the surface of constant energy. As such the probability (in the sense of the frequency interpretation) of finding the representative point in a given volume of the surface of constant energy (i.e. in an observational state of such and such characterisation) is simply proportional to that volume of the hypersurface of constant energy, which is precisely what we would require to justify Gibb's method of virtual ensembles.[29] In other words, there is in the infinite limit, for such systems a clear connection between expected values, in the frequency sense and phase averages calculated using a Gibb's ensemble. Indeed, there is a hierarchy of dynamical systems whose underlying motion is deterministic but which exhibit stronger and stronger stochastic properties. In the strongest of these the so called Bernoulli systems, the deterministic phase flow, at the microscopic level, generates in the sequence of observational states, at the macroscopic level, just that probabilistic independence which one would expect of coin-tosses or the plays of a roulette wheel.

While it is true that the hierarchy of ergodic systems provides a solution to the problem of deterministic probabilities by providing a rigorous frequency interpretation of them the class of systems for which this holds is highly restricted. Few systems are known to be ergodic and many interesting ones for which we require explanation of their statistical properties cannot be ergodic because they satisfy other dynamical constraints on their motion than just conservation of energy and thus their trajectories cannot possibly traverse the whole of the surface of constant energy, so they fail to be ergodic. Further the properties which guarantee interesting statistical behaviour are proved within the ergodic hierarchy only at the infinite limit and it is very far from clear that the finite time averages that we actually observe can be obtained rigorously from the ergodic infinite time limit.[30] Clearly we need to look elsewhere to obtain a more general account of deterministic probabilities.

An interesting suggestion has been made recently by Barry Loewer.[31] He proposes to amend David Lewis's interpretation of probability to yield an account of nontrivial probability in deterministic contexts. Lewis himself, like Popper, though for very different reasons, thought that if the fundamental laws of a theory where deterministic then whatever probabilities it postulated must be understood as subjective degrees of belief. Loewer quotes him to the effect that:

> "To the question of how chance can be reconciled with determinism ... My answer is it can't be done ... there is no chance without chance. If our world is deterministic there is no chance save chances of zero or one. Likewise if our world contains deterministic enclaves, there are no chances in those enclaves."[32]

---

[29] Cf. Gibbs, J. W., *Elementary Principles in Statistical Mechanics Developed with Special Reference to the Rational Foundation of Thermo-dynamics*, 1902 in the edition of Dover Publications, New York, 1960.

[30] An excellent article on the interpretation of probability in ergodic theory is that of van Lith, J., "Ergodic theory, Interpretations of Probability and the Foundations of Statistical Mechanics," *Studies in the History and Philosophy of Modern Physics*, v. 32, (2001), pp. 581-594.

[31] Loewer, B., "Determinism and Chance," footnote 15.

[32] "Determinism and Chance," p. 610.

As is very well known Lewis proposes a "best-fit" account of natural laws. A regularity is a law if and only if it is a theorem of that system of axioms which best balances the two constraints of simplicity and correctness of predictions. Now we can extend this idea to law-like claims about probability. If we were trying to describe the sequence of outcomes of throws of a fair coin we could list the outcomes, but this would result in an enormously complex description (as long as the sequence). However we could simply say that the outcomes formed a Bernoulli trial with probability one-half which is very simple and assigns the actual outcomes to well determined class. Such a probabilistic description very well fits Lewis's description of a law of nature. Now suppose we extend this idea to the history of the Universe as a whole. The theoretical system we have been discussing throughout this section to describe the history of the Universe is actually the combination of Newtonian mechanics, a postulate that initially the Universe was in a very low entropy state (to guarantee that that Boltzmann's statistical explanation of the second law of thermodynamics is true) and that the probability distribution at the origin of the Universe over the initial low entropy state is precisely the standard micro-canonical one: namely that the probability of (now) the Universe being in a particular microscopic dynamical state given that it is an a particular macroscopic state is proportional to the volume of microscopic states which yield that macroscopic state in the phase space surface of constant energy of the Universe. Those three postulates together are a simple best fit theoretical system which captures the facts. This in fact gives a law-like character to initial conditions, but those initial conditions, in particular the third postulate above, play an essential role in the system and thus have a law-like character. The subsequent evolution of the Universe then is entirely deterministic and fixed by the laws together with the actual initial conditions but nevertheless the initial law-like probability distribution over macroscopic states evolves too.

Whether this modified account of physical probability in deterministic contexts is acceptable does depend of course on accepting Lewis's general account of chance and law-likeness and that is a very big issue indeed but at least it shows that a general account is possible and one moreover which employs a fundamental insight of the founder of statistical mechanics Ludwig Boltzmann, that the empirical success of statistical mechanics in explaining all those facts subsumed by the second law of thermodynamics can only be understood by appeal to conditions at the origin of the Universe.

We began traditionally by discussing Laplace's demon and determinism, we should end traditionally by addressing the supposed nightmare scenario of the consequences of physical determinism for human freedom, here too we will conclude that the tradition is very far from the truth.

5. Determinism and Human Freedom

It is interesting to note that a considerable change of attitude took place in the mid-Nineteenth Century with respect to the issue of determinism and human freedom. Prior to that time, especially in the Eighteenth Century, debates about freedom of the will were largely concerned with the very ancient question of whether motives dispose or incline humans to act in the way they do or do they actually necessitate such actions.[33]

[33] There is a very systematic account of these debates to be found in Harris, J, A., *Of Liberty and Necessity: The Free Will Debate in Eighteenth Century British Philosophy*, Oxford University Press, Oxford, 2005. He points out the stark contrast with Nineteenth century debates see especially, pp. 227-235.

Libertarians claimed that no matter how strong a motive to act in a certain way may be, there is always the possibility of an agent having just those motives and acting in a different way. Necessitarians thought otherwise holding that there is a contradiction in supposing that an agent's action may be explained by a motive, that is the motive provides sufficient conditions for the action, while at the same time asserting that that action might not have taken place, might have been different. But the terms of the debate hinge on what might be called the "motive power" of the will. Does it dispose or does it necessitate? By the mid-Nineteenth century however, determinism and the problems it was thought to bring to giving an account of moral responsibility was based upon a different issue, that being the apparent pervasiveness of physical determinism. In his monumental treatise Henry Sidgwick wrote as follows:

> "On the Determinist side there is a cumulative argument of great force. The belief that events are determinately related to the state of things immediately preceding them is now held by all competent thinkers in respect of all kinds of occurrences except human volitions. It has steadily grown both intensively and extensively, both in clearness and certainty of conviction and in universality of application, as the human mind has developed and human experience has been systematised and enlarged. Step by step in successive departments of fact conflicting modes of thought have receded and faded, until at length they have vanished everywhere, except from this mysterious citadel of Will. Everywhere else the belief is so firmly established that some declare it's opposite to be inconceivable: others even maintain that it was always so. Every scientific procedure assumes it: each success of science confirms it. And not only are we finding ever new proofs that events are cognisably determined, but also that the different modes of determination of different kinds of events are fundamentally identical and mutually dependent: and naturally, with the increasing conviction of the essential unity of the cognisable universe, increases the indisposition to allow the exceptional character claimed by Libertarians for the department of human action."[34]

Clearly Sidgwick perceived a nightmare scenario emerging: all scientific theory indicated that the world in its evolution was deterministic and that soon the last citadel of freedom the "will" would fall too to the burgeoning science of psychology which would by extrapolation on all previous unifying developments show the will deterministic in character too. Indeed in fact, the "will" virtually disappeared from psychological theorising very early in the Twentieth Century. Writing some hundred years later Popper again stresses the night-mare quality of a deterministic universe. He wrote:

> "A deterministic physical clockwork mechanism is, above all, completely self contained: in the perfect deterministic physical world, there is simply no room for any outside intervention. Everything that happens in such a world is physically predetermined, including all our movements and therefore all our actions. Thus all our thoughts, feelings, and efforts can have no practical influence upon what happens in the physical world: they are, if not mere illusions, at best superfluous by products ("epiphenomena") of physical events."[35]

---

[34] SIDGWICK, H., *The Methods of Ethics*, Macmillan, London, 1874, 7th Edition, 1907, pp. 62-63.

[35] POPPER, K. R., *Objective Knowledge*, Clarendon Press, Oxford, 1972, p. 219.

And again, he returns to the point, concerning the intolerable nature of the deterministic vision:

> "It is a nightmare because it asserts that the whole world with everything in it is a huge automaton, and that we are nothing but little cog-wheels, or at best sub-automata, within it.
>
> It thus destroys, in particular, the idea of creativity. It reduces to a complete illusion the idea that in preparing [the Compton Memorial] lecture I have used my brain to create something new."[36]

But does the modern analysis of the determinism of physical systems, where it exists, support any of the woeful scenarios mentioned above. I think the answer is no. Indeed what it shows is that the analysis of determinism so far as it can be given with any precision has little if any relevance to the issue of the freedom or otherwise of human action. First there is the obvious point that this has little or anything to do with predictability. One's decisions may be predictable, but still properly construed as appropriate and free. Conversely one might consider the output of some automaton to be determined, perhaps because the output is a function of the input such that identical inputs are followed by identical outputs, as would be required by the account of deterministic physical systems given above, (we might know this for very high level theoretical reasons) but nevertheless it might be the case that the outputs were unpredictable because of some extreme dependency on initial conditions, whose exact values were, even in principle, unknowable by us.

Second there is the very important point that our understanding of determinism is clearly restricted to isolated systems, that is, determinism is only *defined* for such systems. Interactions from "outside" can manifestly destroy the deterministic evolution of a physical system and put bluntly, the very last thing human beings are, are isolated systems. The common response to this point, that we need only consider the evolution of a larger system namely the original system plus the larger environment and think of that as isolated may suffice for some systems, but hardly for human beings. Suppose we took the whole solar system, containing our planet, could we regard that as sufficiently isolated? The truth is we have no idea. Certainly if those scenarios which attribute the extinction of the dinosaurs to a supernovae in another part of the galaxy are correct then we could have two histories of the large system (comprising the solar system) one in which the dinosaurs were not made extinct (by the supernovae) and in which consequently humans did not evolve and the other, the actual one, in which they were and humans did evolve. The larger system (earth plus solar system) would then be indeterministic, we would have two histories identical in their initial stages which then diverged. Perhaps then we should consider the entire galaxy as an isolated system, which it may properly be, for some purposes, but how could we assume that was so for the evolution of life which may very well be dependent very finely on initial conditions. The point is that the problem is not well-defined. We know we can't absolutely isolate systems (think only of the gravitational potential which pervades the entire universe), so we would have to consider the entire Universe as an isolated system. But I have grave doubts as to whether there is a theory of "everything" that can make sense of such a notion.

---

[36] *Objective Knowledge*, p. 222.

Third, we should look closely at the condition which most closely connects the problem of human freedom with determinism, that is, the Principle of Alternative Possibilities. This Principal says that one is free or morally responsible for what one has done only if one could have done otherwise than one did in exactly the same circumstances. But the key issue is what is meant by "exactly the same circumstances." If human beings were like say simple pendulums then it would be easy to say what was meant by "exactly the same circumstances," when the length of the string, the initial displacement angle and the mass of the pendulum were the same. We can say this because we have a complete theory of the motion of simple pendulums. But the same is certainly not the case for human beings. It is even far from clear that there is such a notion for such creatures as "being in exactly the same circumstances." No doubt there is much can be said metaphysically on this issue but from the point of view of natural science there is very little. Clearly what needs to be done is to study very carefully the various theories of empirical psychology of the modules and components involved in human decision making activity to see if they satisfy anything like the constraints we think of as characterising deterministic theories.

Given how we now characterise physical determinism, the question of human freedom understood as an abstract question in general, seems to me rather like the question sometimes posed as to whether human beings are finite Turing machines, or whether the mind is a computer. In one very trivial sense the answer must be yes, since presumably the actions and thoughts of a given human form a finite sequence and all finite sequences are trivially computable by listing them, but that is clearly not what is meant by the question. The real issue is exactly what *is* meant by the question? I suspect the question is simply not well-defined, we don't really know what we are asking when we ask that question even though the notion of Turing computation is perfectly understood, so to with physical determinism and the question of human freedom.

The richness and depth of the modern conception of determinism is very great, the consequences of it for our understanding of Nature have just begun to be explored and it will only be advanced by the careful study of the actual form of the theories that emerge as further questions in physics and psychology are pursued.

6. Bibliography

ARISTOTLE, *De Interpretatione*, English Edition by J. Barnes in *The Complete Works of Aristotle*, The Revised Oxford Translation, Volume 1, Princeton University Press, Princeton, NJ, 1984.

ARISTOTLE, *Physics*, English Edition by J. Barnes in *The Complete Works of Aristotle*, The Revised Oxford Translation, Volume 1, Princeton University Press, Princeton, NJ, 1984.

BELNAP, N., *Facing the Future*, Oxford University Press, Oxford, 2001.

BOHM, D., "A suggested Interpretation of the Quantum Theory in terms of 'Hidden Variables:' I and II," *Physical Review*, v. 85, (1952), pp. 166-193,

BOLTZMANN, L., "Further Studies on the Thermal Equilibrium of Gas Molecules," in BRUSH, S. G. (ed.), *Kinetic Theory*, *v. 2, Irreversible Processes*, Pergamon Press, Oxford, 1966, pp. 88-175.

Butterfield, J., "Determinism and Indeterminism," in: *Routledge Encyclopedia of Philosophy*, v. 3, (1998), pp. 33-38.

Clark, P., "Determinism and Probability in Physics," *Proceedings of The Aristotelian Society,* Suppl. Vol. 62, (1987), pp. 185-210.

Chaitin, G. J., *Algorithmic Information Theory*, Cambridge Tracts in Theoretical Computer Science, Cambridge University Press, Cambridge, 1987.

Church, A. "On the Concept of a Random Sequence," *Bulletin of the American Mathematical Society*, v. 46, (1940), pp. 130-135.

Dewitt, B., "Quantum Mechanics and reality," *Physics Today*, v. 23, (1970), pp. 30-35.

Eagle, A., "Randomness is Unpredictability*," The British Journal for the Philosophy of Science*, v. 56, (2005), pp. 749-790.

Earman, J., *A Primer on Determinism*, The University of Western Ontario Series in the Philosophy of Science, Riedel, Dordrecht, 1986.

Everett III, H., "Relative State Formulation of Quantum Mechanics," *Reviews of Modern Physics*, v. 29, (1957), pp. 454-462.

Gleick, J., *Chaos: Making a New Science*, Viking Press, New York, 1987.

Gibbs, J. W., *Elementary Principles in Statistical Mechanics Developed with Special Reference to the Rational Foundation of Thermo-dynamics*, 1902 in the edition of Dover Publications, New York, 1960.

Fraassen, B. van, "A Formal Approach to the Philosophy of Science," in Colodny, R. G. (ed.), *Paradigms and Paradoxes*, The University of Pittsburgh Press, Pittsburgh, 1972, pp. 303-366.

Harris, J, A., *Of Liberty and Necessity: The Free Will Debate in Eighteenth Century British Philosophy*, Oxford University Press, Oxford, 2005.

Humphreys, P. W., "Is 'Physical Randomness' Just Indeterminism in Disguise?," *PSA 1978*, v. 2, pp. 98-113.

Humphreys, P. W., "Why Propensities cannot be Probabilities," *Philosophical Review*, v. 94, (1985), pp. 557-570.

Humphreys, P. W., "Some Considerations on Conditional Chances," *The British Journal for the Philosophy of Science*, v. 55, (2004), pp. 667-680.

Lewis, D., *On the Plurality of Worlds*, Basil Blackwell, Oxford, 1986.

Lith, J. van, "Ergodic theory, Interpretations of Probability and the Foundations of Statistical Mechanics," *Studies in the History and Philosophy of Modern Physics*, v. 32, n. 4, (2001), pp. 581-594.

Loewer, B., "Determinism and Chance," *Studies in History and Philosophy of Modern Physics*, v. 32, n. 4, (2001), pp. 609-620.

Macfarlane, J., "Future Contingents and Relative Truth," *Philosophical Quarterly*, v. 53, (2003), pp. 321-336.

Martin-Löf, P., "The Definition of a Random Sequence," *Information and Control*, v. 9, (1966), pp. 602-619.

MAXWELL, J. C., "Illustrations of the Dynamical theory of gases." Reprinted in *The Scientific Papers of James Clerk Maxwell*, edited by W. D. Niven, Dover Publications, New York, 1965, pp. 377-409.

MILLER, D., *Critical Rationalism: A Restatement and Defence*, Open Court, La Salle, IL, 1994.

MISES, R. VON, *Probability, Statistics and Truth*, Dover Publications, New York, 1957.

MONTAGUE, R., "Deterministic Theories," in THOMSON, R. H. (ed.), *Formal Philosophy*, Yale University Press, New Haven, 1974, pp. 303-359.

POLLARD, H., *Mathematical Introduction to Celestial Mechanics*, Carus Mathematical Monographs, v. 18, The Mathematical Association of America, Providence, RI, 1976.

POPPER, K. R., *Objective Knowledge*, Clarendon Press, Oxford, 1972.

POPPER, K. R., *Quantum Theory and the Schism in Physics*, Routledge, London, 1992.

POUR-EL, M. and RICHARDS, J., "Non-computability in Analysis and Physics," *Advances in Mathematics*, v. 48, (1983), pp. 44-74.

POUR-EL, M. and RICHARDS, J., *Computability in Analysis and Physics*, Springer-Verlag, Berlin, 1988.

PUTNAM, H., "A Philosopher Looks at Quantum Mechanics (Again)," *The British Journal for the Philosophy of Science*, v. 56, (2005), pp. 615-634.

SALMON, W. C., "Propensities a Discussion Review," *Erkenntnis*, v. 14, (1979), pp. 183-216.

SIDGWICK, H., *The Methods of Ethics*, Macmillan, London, 1874.

SIMON, P., MARQUIS DE LAPLACE, *A Philosophical Essay on Probabilities*, 1820. Translated from the Sixth French edition by F. W. Truscott and F. L. Emory, Dover Publications, New York, 1952.

SKLAR, L., *Physics and Chance*, Cambridge University Press, Cambridge, 1993.

SUPPES, P., *Representation and Invariance of Scientific Structures*, CSLI Publications, Stanford, 2002.

VILLE, J., *Etude Critique de La Notion Collectif*, Gauthier-Villars, Paris, 1939.

WIGGINS, S., *Introduction to Applied Nonlinear Dynamical Systems and Chaos,* Springer-Verlag, Berlin, 1990.

XIA, Z., "The Existence of Non-collision Singularities in Newtonian Systems," *Annals of Mathematics*, v. 135, (1992), pp. 411-468.

Evolutionary Epistemology and the Concept of Life

Franz M. Wuketits

1. Introduction

In his classical paper Campbell wrote: "An evolutionary epistemology would be at minimum an epistemology taking cognizance of and compatible with man's status as a product of biological and social evolution. [...] it is also argued that evolution —even in its biological aspects— is a knowledge process, and that the natural-selection paradigm for such knowledge increments can be generalized to other epistemic activities, such as learning, thought, and science."[1] This includes two implications:

(1) Cognition (including specific types of human knowledge) is a result of evolution by natural selection.

(2) Cognition is an important aspect of evolution and helps us to better understand the evolutionary development of living systems.

Thus, evolutionary epistemology is of relevance for the study of cognition and knowledge as well as for a general conception of life. It is a *naturalized epistemology*, based on the assertion that *knowledge* —traditionally the very subject of epistemology— is a naturally grown phenomenon, so to speak, that can be studied by applying the concepts and methods of the (natural) sciences.[2]

However, two levels or programs of evolutionary epistemology have to be distinguished:[3]

(1) The attempt to account for cognitive mechanisms in living beings by the extension of the biological theory of evolution to those structures that are the (biological) substrates of cognition (brains, nervous systems, sense organs).

(2) The attempt to explain ideas (including scientific theories) in terms of evolution, i. e., to apply evolutionary models to the reconstruction and explanation of ideas.

These two programs are specified in the table below. They are, of course, interrelated in many ways. Only he or she who is willing to accept that cognitive and knowledge phenomena in general are a subject matter of evolutionary thinking, will embrace the evolutionary perspective on specifically human knowledge, including rationality and

---

[1] Campbell, D. T., "Evolutionary Epistemology," in Schilpp, P. A. (ed.), *The Philosophy of Karl Popper*, vol. 2, Open Court, La Salle, IL, 1974, p. 413.

[2] See Bunge, M. and Mahner, M., *Über die Natur der Dinge*, Hirzel, Stuttgart, 2004; Callebaut, W., "Lorenz's Philosophical Naturalism in the Mirror of Contemporary Science Studies," *Ludus Vitalis*, v. 11, n. 20, (2003), pp. 27-55; Oeser, E., *Psychozoikum. Evolution und Mechanismus der menschlichen Erkenntnisfähigkeit*, Parey, Berlin, 1987; and Quine, W. V. O., "Epistemology Naturalized," in Kornblith, H. (ed.), *Naturalizing Epistemology*, The MIT Press, Cambridge, MA, 1994, pp. 15-31.

[3] See Bradie, M., "Assessing Evolutionary Epistemology," *Biology and Philosophy*, v. 1, (1986), pp. 401-450; Oeser, E., *Psychozoikum. Evolution und Mechanismus der menschlichen Erkenntnisfähigkeit*, pp. 9-18; and Wuketits, F. M., "Evolutionary Epistemology: The Nonadaptationist Approach," in Gontier, N., Van Bendegen, J. P. and Aerts, D. (eds.), *Evolutionary Epistemology, Language and Culture*, Kluwer, Dordrecht, 2006, pp. 18-37.

scientific reasoning. But once we are aware of the fact that humans, after all, came up from the ape, we are compelled to take evolution really seriously.

In one single paper it is not possible to present and discuss all aspects and ramifications of evolutionary epistemology and related fields or even to intend to give a broad view of all attempts, past and present, to study the evolution of cognition. My presentation therefore is to be understood only as an invitation to look at cognitive phenomena from an evolutionary point of view and to seriously consider some philosophical implications of such a perspective. I am aware that there are different scientific approaches to the study of cognition (let's just remember the broad and flourishing discipline of *cognitive science*), however, my approach is an evolutionary one, and I still consider evolutionary epistemology as the most promising view of cognition and knowledge.

2. Life and Cognition

It is a (biological) truism that organisms are information-processing systems. They have to act in —and to react to— particular environments; for survival's sake they have to gain some information about what is "out there." In a way, this is true even to plants that react to many changes in their respective environment, e. g. changes in temperature, light intensity, and so on. When advocates of evolutionary epistemology speak of *evolution as a cognition process* or, more precisely, *cognition-gaining process*,[4] then they have in mind that living beings are equipped with information-gaining capacities and that these capacities are results of evolution by natural selection. In this sense, it may even be justified to identify life with cognition.[5] An organism without any ability to recognize some aspects of the outer world would not be able to survive, that is, to reproduce. Any organism is forced to find some resources, to perceive a possible prey, to find shelter in case of danger, and so on and so forth.

Subject of Research	Level of Organization	Aims
Cognitive capacities in living beings and their biological substrates (brain, nervous systems, sense organs)	All levels of animal organization (including humans)	A general biological theory of evolution of cognitive capacities (in nonhuman and human beings)
Human rational knowledge (ideas, scientific theories)	Human mental level	A metatheory of the evolution of human (rational) knowledge

We can lay down here the following basic assertions or, better, tenets of *evolutionary epistemology*:

1. Living systems are information processing systems. Information processing increases their *fitness* and has to be explained in terms of Darwin's theory of natural selection.

[4] See Campbell, D. T., "Evolutionary Epistemology," pp. 413-463; Lorenz, K., *Behind the Mirror*, Harcourt Brace Jovanovich, New York 1977; and Riedl, R., "Evolution and Evolutionary Knowledge: On the Correspondence between Cognitive Order and Nature," in Wuketits, F. M. (ed.), *Concepts and Approaches in Evolutionary Epistemology*, Reidel, Dordrecht, 1984, pp. 35-50.

[5] Cf. Heschl, A., "*L* = *C* A Simple Equation with Astonishing Consequences," *Journal of Theoretical Biology*, v. 145, (1990), pp. 13-40.

2. Animals (including humans) are equipped with particular organs (sense organs, nervous systems or similar structures working analogously [in unicellular animals]) that generate particular "world pictures."
3. The sum-total of information-gaining or information-processing organs, the *ratiomorphic apparatus*, functions in a way similar to a calculation machine; it is analogous to —but not identical with (!)— (human) rational knowledge.
4. Cognitive evolution can be understood as a cycle of *experience* and *expectation*, that means, any organism's ratiomorphic (preconscious) expectation is based on mechanisms that have been stabilized in the course of evolution by experiences made by countless individuals of the respective species over many generations.

Human cognition and knowledge-processing follows basically the same mechanisms. Each of us in his or her everyday life observes some guidelines, so to speak, *innate teaching mechanisms* that were developed and stabilized in the course of our species' evolution. "Everything we know about the material world in which we live derives from our phylogenetically evolved mechanisms for acquiring information."[6] Like other animals, humans are initially not a "clean slate" or *tabula rasa,* but equipped with innate dispositions that already served the purpose of survival in the past. From the beginning on, evolutionary epistemology has contrasted the behaviorists' claim that all types of behavior in any individual organism are just a matter of (individual) learning. Learning is of course an important aspect in the individual life of all "higher" organisms, but it is always based on phylogenetically programmed dispositions that are results of evolution by natural selection. Philosophically, evolutionary epistemology includes an evolutionary interpretation of the Kantian *a priori* as phylogenetical *a posteriori*: All individual *a priori* knowledge rests on phylogenetically developed *a posteriori* knowledge in the respective species.[7] (Evolutionary epistemologists, by the way, do not intend to destroy Kant's remarkable and eminent work but to put it on an evolutionary platform. However, I cannot pursue this aspect further here.)

3. Adaptation

Traditionally, advocates of evolutionary epistemology have fostered an *adaptationist* conception of life and thus an adaptationist explanation of cognitive phenomena. They have followed a most common argument, namely, that evolution means adaptation, that for survival's sake organisms must adapt to their environment(s). As Maynard Smith put it: "We argue from the function to be performed to the kind of organ likely to perform it effectively. Anyone familiar with animals and plants thinks like this a lot of the time, and it is a perfectly legitimate way to think."[8] Indeed, let's just think, for example, of the fins of fish or the wings of birds —they pretty well perform the function of swimming or flying respectively. Hence, it seems obvious that these organs are to be explained as adaptations and that natural

6 Lorenz, K., *Behind the Mirror*, pp. 6-7.

7 See, e. g., Lorenz, K., "Kants Lehre vom Apriorischen im Lichte gegenwärtiger Biologie," *Blätter für Deutsche Philosophie* v. 15, (1941), pp. 94-125; Lorenz, K., "Evolution und Apriori," in Riedl, R. and Wuketits, F. M. (eds.), *Die Evolutionäre Erkenntnistheorie*, Parey, Berlin, 1987, pp. 13-18; and Wuketits, F. M., *Evolutionary Epistemology and Its Implications for Humankind*, State University of New York Press, Albany, NY, 1990.

8 Maynard Smith, J., *The Theory of Evolution*, Penguin Books, Harmondsworth, 1975, p. 17.

selection, generally, favors the better adapted organisms. It does not come as a surprise that, then, evolutionary biologists have always paid particular attention to adaptations. For example, Huxley explicitly stated: "There is a universal process of adaptation."[9]

Likewise, evolutionary epistemologists have argued that the evolution of cognitive mechanisms is mainly a process of adaptation, that our cognitive capacities *fit* our surroundings. Quite famous is the following argument by Lorenz: Just as the hoof of the horse is adapted to the ground of the steppe which it copes with, so our central nervous apparatus for organizing the image of the world is adapted to the real world with which man has to cope.[10] Philosophically, this argument reflects the *correspondence theory* of truth: Our knowledge produces a "true" picture of the external world, what is in our mind corresponds to what is out there. Moreover, prior to rational knowledge our ratiomorphic (preconscious) apparatus already seems to reflect the fundamental structures of the world[11] —otherwise, one could suppose, we would not have survived up to now.

4. Adaptation is not Enough!

Yet adaptation is not the full story of the evolution of cognitive mechanisms, and a *nonadaptationist* approach in evolutionary epistemology has already been developed.[12] Years ago Lewontin critically remarked that the advocates of this epistemology failed "to understand how much of what is 'out there' is the product of what is 'in here'."[13] ("Out there," in an organism's surroundings; "in here," in an organisms' perceiving apparatus.) Let us consider, as an example, the following question: How does a mouse perceive a cat?

From a naive point of view —the point of view of a naive realist, that is— one might suppose that the mouse has a "true picture" of cats and that, for we cannot so easily withstand anthropocentrism, the mouse perceives a cat the same way we humans do. A cat is a *cat*, after all, however, how could we know what a mouse perceives when it is confronted with a cat?! (To be sure, we do not know how it actually is to be a mouse —or a cat, for that matter.) Even if a mouse recognizes just a black spot of any significant size and somehow resembling a cat (or a similar animal), it seems to "know" how to act, what to do —to hide itself or to run away as quickly as possible. Mice and cats are connected by a long evolutionary history, and many millions of mice have made the experience that everything and anything that looks like a cat is something dangerous. Perceiving a cat generates in a mouse a kind of negative feedback, a program in its brain telling *be cautious!* The point is that the mouse does not need to have a "complete picture" of the cat (or even of the group *Felidae*), that its perception does not need to correspond to the cat (anyway, mice do not

[9] Huxley, J., "The Evolutionary Process," in Huxley, J., Hardy, H. C., and Ford, E. B. (eds.), *Evolution as a Process*, Allen and Unwin, London, 1958, p. 2.

[10] Cf. Lorenz, K., "Kants Lehre vom Apriorischen im Lichte gegenwärtiger Biologie," pp. 94-125.

[11] See Riedl, R., "Evolution and Evolutionary Knowledge: On the Correspondence between Cognitive Order and Nature," pp. 35-50.

[12] For details see Wuketits, F. M., "Cognition: A Non-Adaptationist View," *La Nuova Critica*, v. 9/10, (1989), pp. 5-15; Wuketits, F. M., "Evolutionary Epistemology: The Nonadaptationist Approach," pp. 18-37; and the literature cited there.

[13] Lewontin, R. C., "Organism and Environment," in Plotkin, H. C. (ed.), *Learning, Development, and Culture*, Wiley, Chichester, 1982, p. 169.

reflect about "truth"); what is imperative for the mouse is to develop a particular *scheme of reaction* to different objects (including cats) in its environment.

Many traditional biological conceptions give somehow the impression that organisms are *passive* objects shaped by —and adapted to— their surroundings. However, as for example Weiss stated: "Organisms are not puppets operated by environmental strings."[14] Some evolutionary biologists have noticed the failure of a strict adaptationism, among them, prominently, Gould and Lewontin who criticized the adaptationist program as an evolutionary biology of particles and genes rather than constraints and building blocks.[15] They took the spandrels of St. Mark's Cathedral in Venice as an example from architecture: The spandrels exist as a matter of architectural necessity to keep the building up, but, as a by-product, they can also be used for decoration. Analogously, thus Gould and Lewontin argued, many structures in organisms does not necessarily have an immediate function, but are just (non-adaptive!) by-products of the "architectural," organismic constraints. I cannot go into details here, but the message, it seems, is quite obvious: Structures of living beings are not simply determined by the external world but are also, if not primarily, results of "internal constraints."

Organisms are *active systems*, they move, catch prey, build nests, dig holes, and so on, and are therefore not just influenced by their respective environment(s) but have themselves significant effects upon their own surroundings. As *open systems* they maintain themselves by virtue of their systems conditions.[16] They are complex and highly organized systems that contain an enormous number of interacting elements at different levels of their organization. Once again, they are not puppets operated by their environment. Charles Darwin was aware of the meaning of what is now called *organismic constraints* for he stated: "It is preposterous to attribute to mere external conditions, the structure, for instance, of the woodpecker, with its feet, tail, back, and tongue, so admirably adapted to catch insects under the back of trees."[17] Hence, Darwin advocated an "organism-centered" evolutionary biology and was not an adaptationist in a strict sense.[18]

But let's return to evolutionary epistemology. Any organism's perceiving apparatus is not a passive organ simply waiting for getting impressions from outside or being "informed" by objects and processes in the external world. A nonadaptationist approach to understanding cognition includes, at least, the following statements:

1. Cognition is the function of active bio-systems and not of blind machines just responding to the outer world.
2. Cognition is not a mere reaction to the external world but it rather results from complex *interactions* between the organism and its surroundings.

---

[14] WEISS, P., "The Living System: Determinism Stratified," *Studium Generale*, v. 22, (1969), p. 362.

[15] See GOULD, S. J. and LEWONTIN, R. C., "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Program," *Proceedings of The Royal Society, Series B*, v. 205, (1979), pp. 581-598.

[16] See BERTALANFFY, L. VON, *Problems of Life*, Harper and Brothers, New York, 1952; and BERTALANFFY, L. VON, *General System Theory*, Penguin Books, Harmondsworth, 1973.

[17] DARWIN, CH., *On the Origin of Species by Means of Natural Selection*, Murray, London, 1859, p. 28.

[18] See also WUKETITS, F. M., "The Organism's Place in Evolution: Darwin's Views and Contemporary Organismic Theories," in PETERS, D. S. and WEINGARTEN, M. (eds.), *Organisms, Genes and Evolution*, Steiner, Stuttgart, 2000, pp. 83-91.

3. The evolution of cognition is not to be understood as a linear process of step-by-step accumulation of information but a complex process of continuous error elimination.

This brings us to some reflections concerning serious philosophical problems, particularly the problem of *realism*, and to some considerations regarding the concept of life and, finally, to an understanding of human beings.

5. Reality and Realism

As was already mentioned, all organisms are equipped with innate dispositions that are results of evolution by natural selection, this is to say outcomes of selective mechanisms which, among all "initial products," favour and stabilize the one which best copes with the respective conditions of living and surviving. Let's take, as a simple example, hedgehogs. These members of the mammalian order Insectivores rise their spines in order to protect themselves from carnivorous animals like dogs or foxes. This behavior is genetically determined and has been phylogenetically acquired as a comparatively effective survival strategy. Any individual hedgehog raising its spines when confronted wit a dog or a fox follows an old "teaching mechanism" that, like all such phylogenetically acquired mechanisms, contains "information that tells the organism which of the consequences of its behavior must be repeatedly attained and which ought to be avoided in the interest of survival."[19] A hedgehog does not need to have, like a zoologist, profound knowledge of dogs and foxes, it is sufficient that it reacts *adequately* (i. e., for survival's sake) whenever confronted with such objects that *we* name "dogs" and "foxes." A hedgehog is a "hypothetical realist." Based on its innate teaching mechanisms it "expects" that such objects —and everything somehow similar— are dangerous. Its perception does not need to *correspond* (in a strict sense) to the outer world, but just to inform the animal that there is something dangerous around. Advocates of evolutionary epistemology have used the term *hypothetical realism* in this context.

I have used the concept *functional realism*[20] to indicate that

– animals (including humans) are able to identify different objects in their surroundings by the properties these objects seem to have or by the functions they fulfil for the respective animals,
– animals (including humans) learn how to "function" when confronted with different objects in their environment, that means how to react to these objects in order to survive.

In other words: Living beings are generally "realists," but their "realism" is not defined (or even determined) by a kind of metaphysical reality; it is rather the result of a continuos confrontation with concrete objects of the outer world which is guided by the imperative of survival. No organism is forced to recognize "the world," but to perceive those aspects of reality which to "know" is of vital interest. Hence, any species lives in its own *cognitive niche*.[21]

---

19 Lorenz, K., *Evolution and Modification of Behavior*, The University of Chicago Press, Chicago, 1965, p. 16.

20 See Wuketits, F. M., "Functional Realism," in Carsetti, A. (ed.), *Functional Models of Cognition*, Kluwer, Dordrecht, 2000, pp. 27-38.

21 See Uexküll, J. von, *Theoretische Biologie*, Springer, Berlin, 1928.

Organisms do not simply get a “picture” of (parts of) their outer world, but rather develop a particular scheme or way to react to what is happening “out there.” Whenever an animal perceives (sees, hears, smells, etc.) something, it immediately *interprets* what it perceives. My concept of functional realism, then, resembles the notion of *functional coherence*[22] which means that the supposed correspondence or congruence between the objective world and a perceiving subject is to be replaced by a broader conception expressing the close phylogenetic relations between object and subject and showing that they interact as parts of *one* reality.

6. Concept of Life

What I have roughly outlined here is a nonadaptationist view of cognition, the nonadaptationist version of evolutionary epistemology. It is important to note that this, of course, does not mean *anti*-adaptationist. Adaptation *does* play its role in evolution, however, one has to be aware that *adaptability* is not defined by the environment but the organism itself. The environment cannot impinge on organisms arbitrarily any change —a rhinoceros cannot develop into a bird-like creature even if this was of some advantage under particular environmental conditions. Any evolutionary theory and evolutionary epistemology has therefore to pay attention to the organisms’ own abilities to change, its constructional and functional constraints. “Organisms inherit a body form and a style of embryonic development; these impose constraints upon future changes and adaptation. In many cases, evolutionary pathways reflect inherited patterns more than current environmental demands.”[23]

The nonadaptationist version of evolutionary epistemology rests on an *organism-centered* concept of life and is at the same influencing such a concept. Living systems are viewed as hierarchically organized open systems consisting of parts that are mutually related, linked together by feedback loops and regulatory principles. They are multilevel systems; not only its parts or elements determine the whole system of an organism but the organism vice versa determines and constrains the structure of its parts (*downward causation* according to Campbell),[24] so that there is a constant flux of cause in effect in two directions —up and down the levels of any organismic organization. In some more recent publications the importance and meaning of studying the *organismal form* is outlined.[25] The message is that a holistic view, considering the whole organism and its history, is urgently needed for living systems are neither just heaps of molecules (ontological reductionism) nor reflex-chain machines (behaviorism). Irrespective of his commitment to adaptationism, Lorenz correctly stated that “life is an eminently active enterprise aimed at acquiring both a fund of energy and a stock of knowledge, the possession of one being instrumental to the acquisition of the other.”[26]

---

[22] See Oeser, E., *Das Abenteuer der kollektiven Vernunft. Evolution und Involution der Wissenschaft*, Parey, Berlin, 1988.

[23] Gould, S. J., *Hen’s Teeth and Horse’s Toes*, Penguin Books, Harmondsworth, 1983, p. 156.

[24] See Campbell, D. T., “‘Downward Causation’ in Hierarchically Organised Biological Systems,” in Ayala, F. J. and Dobzhansky, T. (eds.), *Studies in the Philosophy of Biology*, Macmillan, London, 1974, pp. 179-186.

[25] See, e. g., Müller, G. B. and Newman, S. A. (eds.), *Origination of the Organismal Form*, The MIT Press, Cambridge, MA, 2003.

[26] Lorenz, K., *Behind the Mirror*, p. 27.

Generally, what one can learn from evolutionary epistemology is that cognition is a crucial aspect in the life of organisms and their evolution. Life means cognition, indeed, for survival —the central biological imperative— of any organism would not be possible without a successful reference to the outer world. However, this "reference" is not to be understood simply as a correspondence between the respective organism and its surroundings. Living beings are active systems in the search for survival and therefore not just "portraying" but also "interpreting" the world around them. Following the biological imperative —survival (!)—, they perceive the world around them according to their own (biological) necessities. They "can explore new environments, feed on different resources, disturb existing ecological balances, destroy the niches of some organisms, and create new niches for others."[27] In particular, this is true to *Homo sapiens*.

7. Concept of Man

Evolutionary epistemology also throws some light on our own nature and helps developing a comprehensive view of man, the *rational animal*, as our species has been traditionally labelled. In the meantime, however, there is no doubt that we humans came up from ape-like creatures and are just one among millions of species that are around on this planet. On the other side, nobody can seriously question some cognitive peculiarities in *Homo sapiens*, especially (according to Tomasello)[28]

- the creation and use of *symbols* (including linguistic symbols, written language, mathematical symbols),
- the creation and use of complex tools and technologies,
- the creation of —and participation in— complex social organizations and institutions.

One can think of other specifically human characteristics, such as intentionality and death-awareness, that other species —as far as we can judge— are not capable of. To what extent other species, particularly our next relatives, the chimpanzees, are capable of tool making, is another problem that cannot be discussed here (however, we should not underestimate the abilities of our brothers and sisters in the animal kingdom!).

The point here is that evolutionary epistemology —as a naturalized epistemology— includes the "natural relativity of mind," this is to say that all mental abilities in humans (language, intentionality, death-awareness, etc.) are to be considered as natural phenomena based on biological structures and functions. In other words: Biological evolution has been the precondition to mental evolution. This is trivial for an evolutionist, from the point of view of traditional —*idealistic*— philosophy (which has still a strong impact especially in the German speaking world), however, it means a heresy. The idea that has predominated Western philosophy ever since Plato "insists that our most distinctive characteristics cannot belong to creatures of flesh and blood: the true me and the true you must be things

27 Hahlweg, K., "A Systems View of Evolution and Evolutionary Epistemology," in Hahlweg, K. and Hooker, C. A. (eds.), *Issues in Evolutionary Epistemology*, State University of New York Press, Albany, NY, 1989, p. 61.

28 See Tomasello, M., "The Human Adaptation for Culture," in Wuketits, F. M. and Antweiler, Ch. (eds.), *Handbook of Evolution*, vol. 1, Wiley-VCH, Weinheim, 2003, pp. 1-23.

essentially incorporeal."[29] Yet from the point of view of evolutionary epistemology this idea is, to say the least, untenable.

As a naturalized epistemology this epistemology follows the basic principles and assertions of naturalism, among them[30]

- everything and anything in this world has its natural origins and has evolved,
- complex systems have grown or developed from less complex systems,
- miracles do not occur,
- there is no extrasensory perception,
- we cannot "transcend" (our own) nature.

Hence, *Homo sapiens* is a natural being, the outcome of evolution by natural selection and in that sense not different from the other species, extinct and extant. All the special characteristics of this species are due to the evolution of its brain, a biological organ that obviously has had an enormous selective advantage.

However, *Homo sapiens* is not sufficiently characterized as a *rational* being. We are capable of *reason*, sure, but we are also carrying our evolutionary history with us, there is still "the ape in us,"[31] that means the sum-total of pre-rational, ratiomorphic teaching masters that guided our phylogenetic ancestors successfully through the world. Lorenz put it as follows: "Instinctive —that is, phylogenetically programmed— norms of behavior which resist all cultural influences play an important part in human social behavior."[32] Let me put it this way: We humans are apes, but we are *also* rational beings; we are rational beings, but we are *also* (still) apes. I am aware that some (humans) might feel taken aback by such conclusions. Yet if we consider our own impact on nature and its devastating consequences[33] we should doubt that rationality has replaced our old teaching masters that evolved and were selected just for the purpose of proximate survival without any look at possible further consequences.

It might seem that, then, an evolutionary epistemologist's concept of man is a rather negative and pessimistic one. But the advocates of this epistemology have not reached a "collective conclusion." Lorenz, for one, concluded that humankind finds itself indeed in a dangerous situation, but that, at the same time "the modes of thought that belong to the realm of natural science have, for the first time in world history, given us the power to ward off the forces that have destroyed all earlier civilizations."[34] I am not that optimistic, however, this is not the place to discuss the possible future of humankind. What does seem clear is that only if we seriously try to reconstruct and to understand the evolutionary roots

[29] FLEW, A., *A Rational Animal and other Philosophical Essays on the Nature of Man*, Clarendon Press, Oxford, 1978, p. 123.

[30] See, e. g., VOLLMER, G., "Was ist Naturalismus?," in KEIL, G. and SCHNÄDELBACH, H. (eds.), *Naturalismus*, Suhrkamp, Frankfurt/Main, 2000, pp. 46-67.

[31] Cf. WUKETITS, F. M., *Der Affe in uns*, Hirzel, Stuttgart, 2001.

[32] LORENZ, K., *Behind the Mirror*, p. 244.

[33] For a most recent review see VERBEEK, B., "The Human Impact," in WUKETITS, F. M. and AYALA, F. J. (eds.), *Handbook of Evolution*, vol. 2, Wiley-VCH, Weinheim, 2005, pp. 243-272.

[34] LORENZ, K., *Behind the Mirror*, p. 245.

of our very nature —including our specific cognitive abilities— should we be able to get some insight into our present situation and to develop some useful ideas for our future.

From the point of view of evolutionary epistemology —particularly in its nonadaptationist version— humans are very complex living systems, and the evolutionary epistemologist's approach to their understanding is a holistic one. It is clear that this approach does not provide comfort to people whose thinking is deeply rooted in idealistic conceptions of humans, but if we are interested in making some progress in philosophy, we can no longer ignore the evolutionist's insight. We are not demigods but part of the biosphere and its complex regulatory principles; we are —more than any other species— actors in the biosphere, but we have to take into account that the biosphere acts back upon our own actions. Thus, if we take evolution seriously we are forced to take nature seriously and to pay tribute to our own *natural history*.

Finally, I should like to stress that the evolutionary world view —including evolutionary epistemology— can be seen as a venerable contribution to a *secular humanism* which is urgently needed in a world in which ideological and religious fundamentalism and obscurantism again begin to darken the minds of many members of our species. But this would require another paper.

8. Bibliography

Bertalanffy, L. von, *Problems of Life*, Harper and Brothers, New York, 1952.

Bertalanffy, L. von, *General System Theory*, Penguin Books, Harmondsworth, 1973.

Bradie, M., "Assessing Evolutionary Epistemology," *Biology and Philosophy*, v. 1, (1986), pp. 401-450.

Bradie, M., "Evolutionary Epistemology as Naturalized Epistemology," in Hahlweg, K. and Hooker, C. A. (eds.), *Issues in Evolutionary Epistemology*, State University of New York Press, Albany, NY, 1989, pp. 393-412.

Bunge, M. and Mahner, M., *Über die Natur der Dinge*, Hirzel, Stuttgart, 2004.

Callebaut, W., "Lorenz's Philosophical Naturalism in the Mirror of Contemporary Science Studies," *Ludus Vitalis*, v. 11, n. 20, (2003), pp. 27-55.

Campbell, D. T., "Evolutionary Epistemology," in Schilpp, P. A. (ed.), *The Philosophy of Karl Popper*, vol. 2, Open Court, La Salle, IL, 1974, pp. 413-463.

Campbell, D. T., "'Downward Causation' in Hierarchically Organised Biological Systems," in Ayala, F. J. and Dobzhansky, T. (eds.), *Studies in the Philosophy of Biology*, Macmillan, London, 1974, pp. 179-186.

Darwin, Ch., *On the Origin of Species by Means of Natural Selection*, Murray, London, 1859.

Flew, A., *A Rational Animal and other Philosophical Essays on the Nature of Man*, Clarendon Press, Oxford, 1978.

Gould, S. J., *Hen's Teeth and Horse's Toes*, Penguin Books, Harmondsworth, 1983.

Gould, S. J. and Lewontin, R. C., "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Program," *Proceedings of The Royal Society, Series B*, v. 205, (1979), pp. 581-598.

HAHLWEG, K., "A Systems View of Evolution and Evolutionary Epistemology," in HAHLWEG, K. and HOOKER, C. A. (eds.), *Issues in Evolutionary Epistemology*, State University of New York Press, Albany, NY, 1989, pp. 45-78.

HESCHL, A., "*L = C* A Simple Equation with Astonishing Consequences," *Journal of Theoretical Biology*, v. 145, (1990), pp. 13-40.

HUXLEY, J., "The Evolutionary Process," in HUXLEY, J., HARDY, H. C., and FORD, E. B. (eds.), *Evolution as a Process*, Allen and Unwin, London, 1958, pp. 1-23.

LEWONTIN, R. C., "Organism and Environment," in PLOTKIN, H. C. (ed.), *Learning, Development, and Culture*, Wiley, Chichester, 1982, pp. 151-170.

LORENZ, K., "Kants Lehre vom Apriorischen im Lichte gegenwärtiger Biologie," *Blätter für Deutsche Philosophie*, v. 15, (1941), pp. 94-125.

LORENZ, K., *Evolution and Modification of Behavior*, The University of Chicago Press, Chicago, 1965.

LORENZ, K., *Behind the Mirror*, Harcourt Brace Jovanovich, New York 1977.

LORENZ, K., "Evolution und Apriori," in RIEDL, R. and WUKETITS, F. M. (eds.), *Die Evolutionäre Erkenntnistheorie*, Parey, Berlin, 1987, pp. 13-18.

MAYNARD SMITH, J., *The Theory of Evolution*, Penguin Books, Harmondsworth, 1975.

MÜLLER, G. B. and NEWMAN, S. A. (eds.), *Origination of the Organismal Form*, The MIT Press, Cambridge, MA, 2003.

OESER, E., *Psychozoikum. Evolution und Mechanismus der menschlichen Erkenntnisfähigkeit*, Parey, Berlin, 1987.

OESER, E., *Das Abenteuer der kollektiven Vernunft. Evolution und Involution der Wissenschaft*, Parey, Berlin, 1988.

QUINE, W. V. O., "Epistemology Naturalized," in KORNBLITH, H. (ed.), *Naturalizing Epistemology*, The MIT Press, Cambridge, MA, 1994, pp. 15-31.

RIEDL, R., "Evolution and Evolutionary Knowledge: On the Correspondence between Cognitive Order and Nature," in WUKETITS, F. M. (ed.), *Concepts and Approaches in Evolutionary Epistemology*, Reidel, Dordrecht, 1984, pp. 35-50.

TOMASELLO, M., "The Human Adaptation for Culture," in WUKETITS, F. M. and ANTWEILER, CH. (eds.), *Handbook of Evolution*, vol. 1, Wiley-VCH, Weinheim, 2003, pp. 1-23.

UEXKÜLL, J. VON, *Theoretische Biologie*, Springer, Berlin, 1928.

VERBEEK, B., "The Human Impact," in WUKETITS, F. M. and AYALA, F. J. (eds.), *Handbook of Evolution*, vol. 2, Wiley-VCH, Weinheim, 2005, pp. 243-272.

VOLLMER, G., "Was ist Naturalismus?," in KEIL, G. and SCHNÄDELBACH, H. (eds.), *Naturalismus*, Suhrkamp, Frankfurt/Main, 2000, pp. 46-67.

WEISS, P., "The Living System: Determinism Stratified," *Studium Generale*, v. 22, (1969), pp. 361-400.

WUKETITS, F. M., "Evolution as a Cognition Process: Towards an Evolutionary Epistemology," *Biology and Philosophy*, v. 1, (1986), pp. 191-206.

WUKETITS, F. M., "Cognition: A Non-Adaptationist View," *La Nuova Critica*, v. 9/10, (1989), pp. 5-15.

WUKETITS, F. M., *Evolutionary Epistemology and Its Implications for Humankind*, State University of New York Press, Albany, NY, 1990.

WUKETITS, F. M., "The Organism's Place in Evolution: Darwin's Views and Contemporary Organismic Theories," in PETERS, D. S. and WEINGARTEN, M. (eds.), *Organisms, Genes and Evolution*, Steiner, Stuttgart, 2000, pp. 83-91.

WUKETITS, F. M., "Functional Realism," in CARSETTI, A. (ed.), *Functional Models of Cognition*, Kluwer, Dordrecht, 2000, pp. 27-38.

WUKETITS, F. M., *Der Affe in uns*, Hirzel, Stuttgart, 2001.

WUKETITS, F. M., "Evolutionary Epistemology: The Nonadaptationist Approach," in GONTIER, N., VAN BENDEGEN, J. P., and AERTS, D. (eds.), *Evolutionary Epistemology, Language and Culture*, Kluwer, Dordrecht, 2006, pp. 18-37.

Conflict between Knowledge and Perception: New Spaces for the Comprehension and Management of the Science around the "New Biology"

Emilio Muñoz

The association between the term "knowledge" and social participation with regards to science and technology is not free from difficulties. The analysis of the level of recognition of society with scientific (and technological) developments reveals a considerable degree of ambiguity and, to a certain point, conflict. This situation is especially evident in the case of developments of the "new" biology, where advances and applications make society face some really conflictive positions, derived from the outcrop of a combination of interests, beliefs and interweaving values, making the process of generation and stabilization of trust in experts extremely complex. All of this refers to the difficulty of associating this process with knowledge.

This situation finds powerful motives in the nature itself of the biological field, where generation of knowledge requires multi and interdisciplinary research and is also driven by a pattern of accelerations and circular paths, making it extremely difficult to adapt traditional tools to the dissemination of knowledge to society (process of acculturation through information and education).

1. Where do we start and Where are we?

Among the terms which have succeeded in describing this society, which wanders between two centuries; the XX or science century, as has been labelled by the historian and academician José Manuel Sánchez Ron,[1] and the XXI century which is still babbling and, therefore still pending a label, is that of "knowledge society."[2] This knowledge society, in principle, appears to be adapted and adjusted to the evolutionary dynamics of developed nations according to the western pattern, although the term does not cease to pose doubts regarding its adequate definition with respect to semantic, political and social references.

From my critical point of view regarding the possible suitability of the term, I believe these shortcomings are in great part due to the lack of effort of definition and meditation

[1] Cf. Sanchez Ron, J. M., *El siglo de la Ciencia*, Taurus, Madrid, 2000.

[2] Javier Echeverría has approached this subject with an experimental metaphor, proposing that the Third Setting is a "territory" where the objective, raw material, to exploit is knowledge; this is why he says that "we can talk about a knowledge society," Echeverria, J., *Los Señores del Aire. Telépolis y el Tercer Entorno,* Ediciones Destino, Barcelona, 1999, p. 79. We can find a review regarding the analysis of scientific knowledge in the book entitled *The Technoscientific Revolution*, which develops the idea of "technoscience," cf. Echeverria, J., *La revolución tecnocientífica,* Fondo de Cultura Económica, Madrid, 2003, pp. 47-48.

In the framework of two European activities (the project "European Comparison of Public Research Systems" —EUPSR— and the EUROPOLIS project financed by the program STRATA), I have reviewed the transition from an industrial society to a service society, based on the growing extent of application of innovation strategies and the emergence of terms such as "risk society," "thoughtful modernization," or "social evaluation," which are behind this process of change describing the knowledge society and aiming to look for the adaptation of public research systems (organizations). Cf. Muñoz, E., "New Socio-Political Environments and the Dynamics of European Public Research Systems," *Working Paper 02-20*, Science, Technology and Society Group (CSIC), *http://www.iesam.csic.es/doctrab2/dt-0220.pdf, 2002*.

on the origins of this term, their basis and how it can be put into practice, that is, how we can understand and develop its evolutionary dynamics in a complex and not always favourable setting.

In our case, this critical concern emerges from questions raised regarding a research programme which can be included under the general title "Study of the social implication and participation in scientific and technological developments," which can also be identified as "Study of the social space of science and technology." For the last twelve years, this research programme has applied a combination of methodological approximations, according to a holistic or, more exactly, a multidisciplinary approach, such as historiography, sociology, although focusing interests in a well defined time frame —the period ranging from the Second World War up to today— all within a relatively well marked geographical area —Spain and its relationship with the European Union— and on the restricted scientific and technological fields of life and biomedical sciences.

The objective of this work is not to present an exhaustive review of the intellectual achievements, nor the problems encountered, in this long decade of socio-political research, although as a necessary introduction regarding the reasons justifying this essay, I do find it appropriate to list some of the research lines undertaken:

– Analysis of the origin and development ("ontogeny") of scientific communities in biomedicine, with a special emphasis in biochemistry, molecular biology, and genetics.
– Studies on Spanish society's public perception (opinion and attitudes) of science and technology, especially with regards to biotechnological applications.
– Analysis and participation in the process of dissemination and communication of science and technology. Thoughts about the concepts of knowledge, information, culture in relation to science and technology.
– Ethical considerations regarding the processes inherent to scientific and technological development (especially related to the production of scientific knowledge and its applications in the field of life sciences) with respect to each of the elements and actors involved in the mentioned research lines: scientific community, civil society and its organisations, political institutions, businesses and media.

2. Some Conclusions from the Available Data

As a result of the intersection of data obtained from these research lines with introspective thoughts, a series of points are established to set the grounds for discussion.

– If we assume that the introduction of the term "knowledge society" has a lot to do with the introduction of *Information and Communication Technologies* (ICT), it seems logical to accept that we are faced with a semantic conflict which limits the concept's functioning. Knowledge is a broad term which includes, as tools for its achievement, information —transmission of something the receptor ignores through a message or system— and communication —action and means through which relationships between individuals are established—, but it does not seem logical that from these tools, however valuable or representative they may be, society as a whole can be codified as knowledge society, as it is a concept which should comprise many more elements and concepts.

– The studies on public perception, opinion and social attitudes towards science and technology in general, and on technological applications closely linked to the concept "knowledge society," have been based for a long decade on the hypothesis of "cognitive deficit," that is, that the acceptance or denial of science and technology is related to the greater or lesser scientific and technological knowledge. However, accumulations of empirical data have questioned the validity of this assumption.[3] The Spanish case, in this respect, is paradigmatic: Spanish citizens, despite their low level of scientific knowledge, usually have a positive attitude towards science and technology, and they support and trust scientists, acknowledging their experience in their own fields. These positive assessments are found even in cases of conflictive applications, such as that of biotechnology applied to agricultural products, a field which in most European countries, members of the EU, encounters critical and distrustful attitudes, if not rejection such as the case of Greece, Italy, Austria or Germany.

On the other hand, recent work by our research group, performed within the framework of the general research programme mentioned previously and with a well-defined objective of a project which aims to study the "governance of science and technology in Spain,"[4] has allowed us, within the Spanish society, to distinguish between processes of information and knowledge, as well as a possible relationship between the conditions of user and consumer.[5] These studies have been carried out with surveys on the subjects of food safety and acceptance of genetically modified food products; Spaniards have sufficient background information to exercise their condition as consumers, but lack the knowledge to enable them to act as users, that is, they are unable to socio-technically value products and applications.[6]

---

[3] This question is a constant concern within the European Union, which appears as a result of the attitudes of sociologists and social psychologists from the intellectual center that is mainly British. They do not resign themselves to accept the data that result from a European reality, I bitterly believe. In this sense, news from CORDIS, the Community R and D Information Service (appearing on the 3rd March 2005 in the service "madrid+d:Noticia") indicates that a research project financed by the European Commission on the public perception of science in 40 countries —a meta-analysis, no doubt— concludes that, although a link has been established between the levels of science knowledge by the public and their support on the matter, the question of popular support of science needs more than just an additional scientific education.

The work coordinated by Nick Allum from Surrey University (Guilford, United Kingdom) has correlated the results from approximately 200 surveys carried out between 1998 and 2003 in several countries throughout the world. The team in charge of the study notes that an individual's scientific knowledge is simply one more of the numerous factors that explain attitudes toward science, as there are other important factors such as moral values, religious beliefs and political trends.

We obtained similar results in our analysis and different publications: Muñoz, E., "Percepción pública y Biotecnología. Patrón de conflicto entre información, conocimiento e intereses," in Iañez Pareja, E. (ed.), *Plantas transgénicas: De la Ciencia al Derecho*, Editorial Comares, Granada, 2002, pp. 124 and 129-137. Muñoz, E., "Los problemas en el análisis de percepción pública de la Biotecnología: Europa y sus contradicciones," in Rubia Vila, F. J., Fuentes Julian, I. and Casado de Otaola, S. (eds.), *Percepción social de la Ciencia,* Academia Europa de Ciencias y Artes and UNED Ediciones, Madrid, 2004, pp. 159-162.

[4] The referred project is entitled "Civil society and the management of science and technology in Spain" (FECYT-CSIC agreement, annually approved).

[5] Cf. Muñoz, E., Plaza, M., Ponce, G., Santos, D. and Todt, O., "Alimentos transgénicos y seguridad. Información y conocimiento en la opinión de los consumidores españoles sobre alimentos transgénicos," *Revista Internacional de Sociología* (RIS), v. 41, May-August, (2005), pp. 93-108.

[6] As a suggestion and desire to bring about discussion, I dare say that the situation is different in the case of ICT (Information and Communication Technologies) where citizens act both as consumers and users thanks to the access, proximity, familiarity and dissemination of these technologies.

– Studies regarding the characteristics of dissemination and communication (to the public) of science and technology also reveal interesting results which allow the following conclusions to be drawn:

a) The sections or subsections pertaining to scientific and technological knowledge related to life sciences and biomedical sciences (such as health, environment, and food safety) enjoy considerable prestige in written mass media, generally covered under the headlines of the "Society" section. However, the events related to ICT, which a decade ago showed a similar pattern of dissemination as that which I have just described, have dissappeared from this section of the newspaper to be moved, in regular newspapers, to the sections on "Economy" or to become a significant reference in specialised economic journals. On the other hand, subjects related to life sciences, biomedical sciences, and their applications in health, environmental, and agricultural food sectors are frequently included in newspaper supplements or sections devoted to science and technology, whereas the uses of ICT are essentially used as indicators of the degree of socio-economic development of a country or society, and are therefore found in newspapers as a product of social research.

b) Scientific journalists in Spain have achieved a reasonable level of quality in such a way that, results from opinion and attitude surveys indicate that a greater recognition is found in this type of journalism than that shown towards journalists in general, at the same time as Spaniards grant them a certain degree of trust in treating matters related to science and technology.[7]

c) Analyzing the contents on subjects related to life sciences and biomedicine, i.e. agricultural biotechnology, cloning, gene therapy, xenotransplants, reveals the existence of an acceptable degree of rigour in treating this information, as well as the importance of interests in the approach and treatment of these subjects. If a subject arises for political reasons, the information is adjusted to those objectives; if, on the other hand, it emerges due to scientific and technological interests —generally because of international influence and specific motivation of the journalists involved, the information is treated according to these positions. If a subject arises due to a combination of interests, the topic is usually high priority information, and its treatment in mass media follows specific guidelines which can even lead to the headline appearing on the cover.[8]

Another interesting fact, although subject to a more in depth analysis, is that a greater number of sources are consulted when a subject has political and social consequences than when it pertains strictly to the scientific and technological fields.

---

[7] Cf. Muñoz, E. and Plaza, M., "Imágenes de la Ciencia y la Tecnología en España a través del espejo de la encuesta de percepción 2004," *Percepción social de la Ciencia y la Tecnología en España-2005,"* FECYT, Madrid, 2005, pp. 135-161.

[8] For detailed information on these analysis, see volume "Opinión Pública y Biotecnología", *Sistema,* n. 179-180, March (2004), in particular the articles by Muñoz, E., "Opinión pública y Biotecnología: Un 'puzzle' con muchas y variadas piezas," pp. 3-13; Diaz Martinez, J. A. and Lopez Pelaez, A., "Biotecnología, periodismo científico y opinión pública: Consenso, disenso, evaluación democrática y difusión de los avances tecnológicos en el siglo XXI," pp. 135-158; Plaza, M., "Análisis de contenido sobre el tratamiento de las aplicaciones biotecnológicas en la prensa española," pp. 171-186: and Santos Benito, D. and Dordoni, P., "Opinión pública y debate ético-social sobre un reto de la Biotecnología: Los xenotrasplantes," pp. 187-205.

Next, I find it pertinent to refer to the dissemination practice I have been involved in for the last two decades, especially in the last five years. In the development of this activity, I have been able to observe the differences between scientific and technological subjects with regards to their interest, access and assumption by the actors involved in the process of dissemination (circulation) of science and technology. In general, the subjects closest to us, with greater ease for assumptions and assimilations, are those related to mechanical and productive technologies, those which affect public health and those concerning the description, conservation and preservation of natural resources. However, the concepts related to basic biology developments, derived from biochemical, molecular and cell biology, immunology or molecular genetic approaches, are more remote from non-expert citizens and, therefore, are not very open to their communication (both due to their own nature, and their rapid development and transformation, as well as the fact that they are written in their own language, which is almost encrypted).

The approach, close to the thematic field "Science and Society" which is the basis of our previously mentioned research programme on social participation in the development of science and technology, tries to understand the public's needs in order to preserve their trust and facilitate their intervention in the management of that development dynamics. For this, information regarding ethical matters must be made more accesible, the dialogue on ethics with non-government organisations, industry, religions and cultural communities must be promoted as well as the active participation of researchers themselves[9] as a result of their awareness regarding the value of the ethical aspects of their activities.

It is an indisputable reality that rapid scientific and technological progress has brought about a series of relevant ethical questions which must be taken into account when evaluating advances in science and technology. These questions are especially applicable in the field of biology, in particular in what we can define as "new biology," with its impacts on medicine and attainment of ever greater challenges in the basic research area. From all of this, in the last decades, the emergence and vibrant development of "bioethics" has taken place and has expanded its influence on any aspect covered by research and technological development in life sciences. It is important to note the problems this success has generated, especially in North America, United States and Canada. Bioethics, which was conceived as an instrument for meditation, debate and moral assessment of advances in biology and biomedicine, and configured as a space to contribute to the dialogue between the different actors involved in the scientific-technological progress and its applications, and generate, in this way, shared trust, has seen a progressive deviation from its objectives, as a victim of its spectacular success.

The advances derived from the new biology, genetic engineering, gene therapy, cloning, repair medicine, information on genomes, have been running into ethical frontiers but, at the same time, have served to awaken and inspire companies' opportunism, as these advances have served as a basis for the embellishment of capitalism: risk capital investments, speculative capital —stock market—, new business initiatives. This context has generated a

---

[9] This orientation fits at all with the activities of the "Action Plan on Science and Society", approved by the European Commission on December 2001, as a basis for the activities developed under this frame in the *VI Framework Program*, aiming to contribute the responsible applications of the scientific and technological progress with respect to basic ethical values.

whole epiphenomenology which can be summarised by the existence, mainly in the United States, of important connections between specialists in bioethics (bioethicists) and industry. This situation is accurately reflected in an article published in *Nature Biotechnology*.[10] This tendency of (North American) bioethics, which can be described as managerial, has begun to bring about criticism. Carl Elliot, cited in an article by Leigh Turner, as one of the most qualified spokesmen of these criticisms, has described the situation as very dangerous, as bioethics has become a hostage to corporative interests.[11]

The situation in Spain is different, as the debate on bioethics is not centered around the area of economic or industrial interests, but on the field of religious beliefs.

In any case, we can conclude, perhaps in an excessively schematic way, that bioethics, despite the hopes raised of it becoming a bridge between experts and society in the assessment of scientific progress in life sciences, does not appear to be able to achieve that goal, in light of current circumstances. In spite of this —and despite this discouraging conclusion— I believe that the introduction of ethical perspectives in communication processes of scientific-technological advances, especially in the field of biotechnology, is a good instrument to increase trust between the person providing information and the one receiving it; emphasizing that trust is essential for a non-expert society.[12] I have developed this thesis in presentations carried out within the framework of two European activities; one promoted by the Task Group on Public Perceptions of the European Federation of Biotechnology[13] and the other within the Europabio Convention, both throughout the year 2003.[14]

In these frameworks, I have proposed that communication ethics is an attitude towards a technological activity which must involve all actors which play a role in the development of technology, biotechnology in our case. It is professional ethics to take into account moral beliefs of the professions, for which the term "interethics" has been coined, as this ethics comprises various professions: producers of knowledge, translators (journalists and mass media), social actors and businessmen.

3. Biology: Novelties and Complexity

In the following part I will try to develop some arguments which will help us understand the difficulties encountered by common citizens, ignorants in the matter, to understand the advances and developments in biology and, consequently, express their opinions, with the objective of them contributing in the management of science and technology in the areas of life sciences and biomedical science.

10 Cf. Turner, L., "Bioethi$ Inc," *Nature Biotechnology*, v. 22, n. 8, (2004), pp. 947-948.

11 The opportunism of polarization in genetics' ethics ("genethics," in English) has also been the target of tough criticism by Leigh Turner. Cf. Turner, L., "The Tyranny of 'Genethics'," *Nature Biotechnology,* v. 21, n. 11, (2003), p. 1282.

12 Cf. Eaton, M. L., *Ethics and the Business of Bioscience*, Stanford University Press, Stanford, 2004. This book is reviewed in Dhanda, R. K., "Making the Case for Ethics," *Nature Biotechnology*, v. 22, n. 8, (2004), p. 951.

13 Cf. Muñoz, E., "Ethical Dimensions in Biotech Communications," *EU Focus Workshop on "Who should communicate with the public and how?,"* Task Group on Public Perceptions, European Federation of Biotechnology, Madrid 13-14 June, 2003.

14 Cf. Muñoz, E., "To Gain Trust: Ethics and Communication in Biotechnology," *Cordia Europabio Convention 2003,* Vienna, 2-4 December, 2003.

This discussion is based on two concepts: firstly, the idea of diversification, of junction in the development of science, in line with what Michael Serres and his school have tried to write in the history of science,[15] an idea which assume to evolve in the same line of thought as that which I have used in my approach to explain the development of biotechnology in some publications;[16] secondly, as a member of a group that works in the relationship between science, technology and society, I think it is appropriate to tackle the concept of "dialectical biology," in an extension of the analysis proposed by Lewins and Lewontin in their book *The Dialectical Biologist*[17] which, in their case, they adopt to consider the political implications scientific practice has, and in our case we apply to the relationships between the whole and the parts in living organisms, and the interactions organisms have between their same species and between other living organisms.

An essential reference to carry out this analysis is offered by the *Encyclopedia Britannica*.[18] The way in which it handles what is understood by biology, and its consequence in the "biological sciences," is, I believe, a basic starting point to put forward our proposal. The following lines are written with this objective.

Lastly, analyzing the evolution of biological sciences reveals a close relationship between the advances, forking processes and emergence of new disciplines, or different approaches in studying the different levels of organization of life (its fundamental units) with the development and use of new instruments. Among these we can mention: different resolution levels from the cell to the atom; techniques for compound separation based on size, charge, amphipathic nature or all of these parameters combined; application of the centrifugal force to separate or analyse macromolecular components, such as nucleic acids and proteins; techniques for fragmenting these macromolecules and combining them with the characterization of their fragments to establish sequences (sequencing techniques); spectroscopic techniques with different emission and identification sources from radioactive tagging to fluorescent labelling; arrays, similar to those used in electronics to study, in the broadest way possible, the functional image of cells or organisms; computational techniques to exploit sequence data and identify analogies and divergence of molecular components. In this context, of extensive collaboration between the search for knowledge and the use of techniques happily paired between scientific research and technology, we can apply a "techno-scientific" nature to the field of biological research.[19]

4. Biology and Biological Sciences

In its general section, the *Encyclopedia Britannica* defines *biology* as "the study of living creatures, organisms and vital processes." Next, it establishes the divisions

[15] Cf. Serres, M. (ed.), *Elements d'histoire des sciences,* Bordas Cultures, Paris (printed in Poitiers), 1989.

[16] Cf. Muñoz, E., *Biotecnología, Industria y Sociedad: El caso español,* Fundación CEFI, Gabinete de Biotecnología (GABIOTEC), Madrid, 1997.

[17] Cf. Lewins, R. and Lewontin, R., *The Dialectical Biologist,* Harvard University Press, Cambridge, MA, and London, England, 1990.

[18] Cf. Ordoñez, J., Navarro, V. and Sanchez Ron, J. M., *Historia de la Ciencia,* Espasa-Calpe/Austral, Madrid, 2003; see also Serres, M. (ed.), *Elements d'histoire des sciences,* passim and The New Encyclopedia Britannica, *The Biological Sciences,* v. 14, Fifteenth Edition, 1990, pp. 920-977.

[19] Cf. Echeverria, J., *La revolución tecnocientífica,* passim.

according to disciplines or the methodological approaches which distinguish a) between the object under study, *zoology* which is oriented towards the study of animals and *botanics* directed towards the study of plants; b) between the level of resolution: *anatomy*, study of organisms' structure; *morphology*, study of the anatomical structures related with their function; *physiology*, exploration of the way organs work and the physical and chemical processes of living organisms; *cytology*, study of the cells at a microscopic level; *embriology*, study of fetal development; *ecology*, study of organisms and their interaction between other organisms and their environment; *genetics*, search of the mechanisms of heredity; *biochemistry*, study of the chemical processes of living organisms —here we can observe a certain degree of overlap with the definition of physiology— therefore, in my opinion, the following should be added to this definition —"starting from the separation and characterization of its basic components, lipids, proteins, carbohydrates and nucleic acids"—; *molecular biology*, study of vital processes at a molecular level, a definition which is not incorrect but which in my opinion is incomplete and not completely explanatory.

This first incursion in the field of characterization of biology from the *Encyclopedia Britannica* sources reveals the existing differences in the intellectual framework which maintains the definitions of the disciplines or approaches in the study of organisms according to their degree of novelty; a greater age corresponds to a greater stability and clarity in the definitions.

5. Specialization and Biological Sciences

This debt begins to be paid off by the source itself when the *Encyclopedia Britannica* tackles the subject from the idea of "biological sciences" and with the in depth treatment allowed in the *Macropaedia* sections of that work. In this analysis, biology clearly appears as a field of research where a large number of disciplines come together. We can conclude that research in biological sciences is a result of hybridization of great principles from other sciences, such as physics and chemistry, and in this way, a junction, diversification is produced with the emergence of new disciplines such as biochemistry, biophysics and molecular biology.

The considerable broadness of biology as object under study, promotes its classification into different branches of knowledge but, it is important to insist that, despite apparent differences, all subdivisions, whether they are based on discipline or strategy, are closely related. In this way, although it was common practice for significant periods of biological research developments to separate the study of plants (botanics) from that of animals (zoology), and the study of the structure of organisms (anatomy and morphology) from the study of their function (physiology), the current trend is to study all biological phenomena and processes shared by all living organisms.

Research in biological sciences aims to analyse the activity of living organisms from the perspective of the physical-chemical processes supporting such activity. The most frequent approach is to carry out this inquiry starting at the different levels which correspond to life units. The molecular level, thanks to the development and use of tecniques and tools with ever greater resolution, allows us to learn not only about the structural organization based on physical-chemical principles which are invisible, but also about how live matter is able to reproduce itself in molecular terms.

Cell biology, whose object of study is the cell —basic unit of living organisms in structural and functional terms— and whose origins date back to the XVII century with the discovery of the microscope, has reached maturity with the advent of new instruments, both at a microscopic resolution level and at separation and individualized study of the cells. Before the cellular approach, organisms where studied as a whole ("organismic biology"), a research area which is still considered relevant in terms of biological organization. Population biology is concerned with research on groups or populations of organisms inhabiting a certain geographical area or region. Within this organizational level, studies on the role of certain plants or animals in the complex and self-regulated inter-relationships which exist between the living and inert world are included, as well as studies on the production and construction of controls which maintain these relationships in a natural way. From here, ecology and population genetics have arisen as a result of the diversification process of disciplines and approaches in the search of knowledge mentioned as a conceptual basis of our discourse.

The other field of biological study according to the object of study, plants (*botanics*) or animals (*zoology*), has also experienced a diversification process based on the same principle: that is, the type of organism under study; in this way, we have *ornithology*, study of birds; *icthiology*, study of fish; *micology*, study of fungi; *microbiology*, study of microorganisms; *protozoology*, study of mobile unicellular creatures (*protozoa*); *herpetology*, study of amphibia and reptiles; *entomology*, study of insects; and *physical anthropology*, study of man.

6. From Diversity to Homogeneity and Start Over

The course of development in a loop followed by knowledge in biology and medicine can be explored with the concurrence of a historical analysis up until today, where biology has reached an exponential growth and a growing relevance in the society in which we live in. Science history reveals the existence of special moments and circumstances where great advances have taken place. These jumps are usually the reflection of the coincidence of two factors: the presence of a creative mind, able to go against established ideas and propose new hypothesis and the availibility of technological instruments to test such hypothesis.

6.1. Antiquity. The Greek World

As we have seen, biology has two lines or fields of action with different historical developments. The line based on the study of organisms dates back to antiquity. Although it is not known precisely when the study of biology began, it is thought that primitive man must have developed understanding about the animals and plants which surrounded him. In fact, his survival depended on his ability to identify non-poisonous plants he could eat as well as knowing the habits of dangerous predator animals. Therefore, it seems likely that the history of biological knowledge precedes the time when the first registers were produced. These registers coincide with the bas-relief of the Assyrian and Babylonian people who collected plants for growing and for veterinary medicine practices.

It has been confirmed, from papyrus and utensils found in tombs and pyramids, that the Egyptians had knowledge about the medicinal value of certain plants 2000 years before Christ. Even in previous periods (2.800 years B.C.), Chinese knew about the therapeutic properties of a considerable number of plants and they had discovered others, such as soya, used in food products. A well developed agriculture was present in India about 2.500 years B.C. Plants not only served as a source of food, but a document dating back to the VI century B.C. described the use of almost a thousand (960) medicinal plants and also included information on medical subjects such as anatomy, physiology, pathology and obstetrics.

All the biological knowledge collected by the Babylonians, Assyrians, Egyptians, Chinese and Indians was directed at understanding the supernatural more than trying to interpret what was natural, as they lived in a cultural environment dominated by demons and spirits within an unpredictable sphere. Greek civilization introduced an important turn in this mystical course. Approximately 600 years B.C., a Greek school of thought recognized the concept of causality which had a profound influence on scientific research. This group assumed the existence of a "natural law" governing the universe which could be subject to trial by observation and reduction.

In the Greek-Roman world, a series of figures of thought contributed to the formulation of theories on man, the origin of life and even the first proposals of an evolutionary theory. Thales of Mileto and his disciple Anaximander, for example, postulated several origins for life and the main elements supporting it, from water as a first material to the formulation of four elements: land, air, fire and water. Anaximander proposed the first evolutionary theory, considering that life emerged spontaneously from mud and the first animals were fish, which abandoned water to live on dry land where they gave rise to the first animals via a "transmutation" process.

The search for knowledge in biology continued to be present in ancient Greek and Rome with the pre-eminent figure of Aristotle as a reference. Aristotle covered all fields of knowledge, including biology. He established a first classification of living creatures and studied the problems of reproduction and heredity. He believed species were not stable entities, although he did not include evolutionary ideas, and he conferred great importance to teleological causes; he also undertook studies in comparative anatomy. Although botanics is not the field in which Aristotle has been given greatest credit, he wrote two treaties on plants, describing almost 500 species. In this task, he was followed by Teofrast who essentially undertook a descriptive work, making no effort to order and classify.

At the same time, the development of medicine was taking place with Hippocrates playing a major part. The library and museum of Alexandria was created, with Herophilus standing out in these tasks. Another important figure in the field of medical practice was Galen of Pergamum who, due to his connection with gladiators begins to learn about the problems of wounds and their relationship to structures which he can study in depth from his experience.

6.2. The Arabs and Biology

The decline of science in the Greek world begins with the disappearance of Aristotle and Teofrast. From the III to the XI century, biology is a field ruled by Arabs. By themselves they were not especially innovative, but they were scholars, recovering the works and knowledge of Aristotle and Galen of Pergamum, among others, to become guardians of biological knowledge. Among Arab biologists it is worth mentioning al-Jahiz with his book Kitab al-hayaivan ("Book of animals") where the togetherness of nature was emphasized and the relationship between different types of organisms was recognized. Al-Jahiz believed land was made up of masculine and femenine elements and, therefore, assumed the Greek doctrine of spontaneous generation to be a reasonable idea. Ibn Sina, or Avicenna as he is popularly known, was a Persian scientist who lived in the beginning of the XI century and who is considered to be Aristotle's successor, decisively contributing to the assimilation and spreading of biological knowledge.

6.3. Development of Botanics and Zoology

During the XII century, when botanics experienced an important growth and recognition thanks to the interest in medicinal plants, the development of biology was sporadic. Arabian science has a pre-eminent position with respect to Latin, Byzantine and Chinese science, although its decline was starting to show. Latin culture begins an important take-off, as revealed by the figure of Albert the Great, a wise German of the mid XIII century, who is considered the most important Middle Ages naturalist. His treaties on plants and animals were based on scientific authorities from Greek civilization, mainly Aristotle. His works contained, however, original observations and data, such as contributions on the anatomical study of leaves and plant innervations. His skeptical position regarding the superstitions present at the time and the re-introduction of Aristotelian biology had profound effects in the future of European science.

Throughout the Middle Ages, Italy became a centre of reference for science, especially in agriculture and medicine. One of the most important advances was observed in anatomy, with the introduction of dissection in medical studies. It is worth mentioning Mondino dei Liuccí, the most famous anatomy scholar who developed his work at the beginning of the XIV century.

6.4. The Renaissance

It is interesting to note the important revival of Renaissance biology thanks to the participation of actors from other fields of knowledge. Painters decisively contributed to the anatomical knowledge of the human body. The introduction of the printing press was another factor which had a profound effect in the middle of the XVI century, together with the improvement of woodcut art, which linked to the greater availability of paper lead to changes in the management of illustrations which could be printed in large quantities, contributing in this way to their circulation. The diversity of animal groups was a point for meditation which gave rise to recognizing the importance of the concept of homology to establish groups in different units.

The XVI century was witness to an extraordinary development in botanics which did not occur in zoology. This biological field tried to follow a similar path to botanics where illustration of texts by renowned painters was a decisive factor for identifying and characterizing species. However, the description of animals lacked precision, and together with imaginary representations redounded in the very limited scientific value of zoology texts, independently of their artistic value.

Advances in anatomy which, as previously mentioned, showed a combination of humanistic knowledge did not produce a response in the contemporary scientific field. For this reason, the origins of anatomy are attributed to André Vesali, a Belgian doctor who studied in the outstandingly traditional faculties of Louvain and Paris, where he became familiar with the work of Galen of Pergamum. Despite his reputation as a professor, due to disagreements with his superiors, he moved to Padua in 1537 where he began an extensive reform programme both in education and research in anatomy: he abolished the practice of dissections performed by assisstants, so Vesali carried them out personally, which gave a greater closeness to the knowledge and teaching of anatomy.

The Spaniard Miguel Servet was a disciple of Vesali. He was a doctor and theologian who discovered the pulmonary circulation of blood flowing from the right hand side to the lungs, proving in this way, against what was believed up until then, that blood does not go through the central wall of the heart.

6.5. Expansive Period

Sixteen centuries of history had not given biology the statute of science achieved by mathematics, physics and even chemistry. From the XVII century its expansion begins through a series of scientific, technical, organizational and strategic advances.

It is an almost impossible task to describe the path of biology in the last four centuries, even with the brief and schematic way used up to know. We think it is more feasible and relevant to offer a series of flashes to give an idea of the richness and diversity of the achievements and, therefore, the complexity in trying to cover and understand the current situation of biological knowledge.

Century	Foundations and Principles	Main Achievements (Actors)
XVII	Systematization	Blood circulation (W. Harvey)
	Classification	
	Socialization	Scientific societies
	Development of microscopy	Tissue structure (animals and plants) “Animalculus,” “small organisms” (A. van Leeuwehoek) Discovery of the cells (R. Hooke)
XVIII	Taxonomic principles	Linneo
	Spontaneous creation	Buffon / Needham
	Evolutionism	Lamarck
	Paleontology	Cuvier
XIX	Expeditions	Botanics
	Experimentation	End of the spontaneous generation (Pasteur)
	Link between botanics and zoology	Cellular Theory (Schleiden / Schwann)
	Darwin	Theory of evolution
	Modern embriology	Fertilization / chromosomes
	Heredity	Mendel

6.6. Modern Biology: From Parts to the Whole

Modern biology has tried to move forward in understanding the characteristics of living organisms in terms of the physical and chemical properties of their components. This is what has led to hybrid disciplines: biochemistry, whose aim is to identify and chracterize the structure and function of the basic components of organisms —lipids, carbohydrates, nucleic acids and proteins—; biophysics, which tries to understand not only the electrical properties which allow the nervous system to work, but the thermodynamic and quantum interactions and principles which help us understand how biological cells, as well as their components self-assemble and form correctly functioning "quasi-cybernetic" systems.

Molecular biology, and its very close relative, molecular genetics, arise from the junction of these positions. The original objective of molecular biology was to study, through x-ray diffraction, the functionally most significant macromolecules, proteins and subsequently nucleic acids, of living organisms. Molecular genetics aimed to explain the processes of heredity in terms of discrete elements, genes and chromosomes.

From the proposed structure of deoxyribonucleic acid (DNA) by Watson and Crick in 1953, there was a common purpose in the objectives of molecular biology and molecular genetics; trying to understand the phenomena of gene expression, the functionality of nucleic acids (role of ribonucleic acid, RNA) in the processes of replication, transcription and translation leading to protein synthesis, and their role in regulatory processes both from a self-regulatory and from an extrinsic regulation (effect on DNA and RNA) viewpoint. Throughout this process of knowledge production in biology the concept of reductionism was coined and developed, a concept assigned to the efforts to explain complicated phenomena in terms of simpler concepts, used profusely by scientists in biology. The culminating point of this position is reflected in the phrase by Monod "what is valid for *E. coli* bacteria is valid for an elephant."

This reductionist concept always poses problems to humanists and sociologists who are used to a more holistic approach. Criticism has emerged in recent times even among scientists themselves (i.e. the work published in 2004 by Mattick)[20] regarding the predominance of a molecular vision when interpreting biological phenomena which has been exacerbated by the scientific, social and political hyperbole surrounding the "Genome" project. This project has placed modern biology in the signt of media attention, which has lead to polarizations by citizens regarding the "genetic reductionism (determinism)," when assuming the idea that a single (and specific) gene can be responsible for a certain facet of human behaviour (homosexuality gene, criminality gene, risky behaviour gene).

It is odd that the assumptions made by citizens regarding the possible deterministic role of genes, considered individually and in an isolated manner, disagree with what we learn from the in depth study of genomes. As more information becomes available on sequences from different genomes, we learn about the great similarities in gene endowment of very different organisms, from bacteria, going through the vinegar fly, the worm, or even yeast, up to man. This notion leads to the conclusion that a genetic basis is needed for the dynamics of living beings, mainly for their build up and development, but it is not enough.

[20] Cf. MATTICK, J. S., "Introns," *Investigación y Ciencia,* n. 339, (December 2004), pp. 26-33.

There are two main principles which contribute to gene modulation and activity: regulation and evolution, which in addition to their instrumental and theoretical value they serve to point out the importance of the environment in the development of life. These two concepts combining theoretic value with utility are not easily understood by non-expert citizens in general, nor in particular by mass media and their main actors, journalists.

Despite these difficulties, it is important to emphasize that these principles have been and are basic in the philosophical discourse on biology. The great evolutionary biologist, as well as relevant philosopher and historian in biology, Ernst Mayr,[21] and Richard Levins and Richard Lewontin in their book *The Dialectical Biologist*, a title which contains a subtle metatphor, have questioned reductionism from the force of the dialectical method, combined with inherent elements to the evolutionary theory: continuous change and co-determination between organisms and environment, the part and the whole, structure and processes, science and politics.

7. A New Turn: Molecular Biology Retrieves Biology of System

The new systematic approach aims to integrate and analyse various tendencies of biochemical information in a way which is not easily understood by human intelligence so theories, raised as working thesis, can be validated without the need of spending years on trial and error assays.

The term "biology of systems," whose origins date back to the XIX century in relation to the knowledge on embriology and mathematic analysis of networks, has been applied with a historical perspective to the analysis of various areas of biology, including ecology, developmental biology and immunology. At this time, the genomic revolution has catapulted molecular biology to the field of biology of systems, bringing back the physiological sense to studies on organisms. The systematic approach applied to studies on unicellular organisms and well defined cell lines from superior organisms, is allowing ultimate steps to be taken for understanding the basis of scientific and biotechnological applications.

Contemporary biology of systems arises from the junction of two research lines in molecular biology: a first one based on seminal works on nature and genetic material, the structural characterization of macromolecules and subsequent advances in recombination and high resolution technologies; the second one, which has a more distant, but no less robust for the concept of biology of systems, basis than traditional molecular biology, is based on the thermodynamic theory of irreversible (far from equilibrium) processes, on the determination of biochemical paths and the acknowledgment of feedback controls in unicellular organisms, as well as on the increasing discovery of existing networks in biology.

8. Aspects to be Further Considered

This dialectical dynamics is quite in agreement with what biological science means in its essence and dynamics. Our hypothesis is that its development has been and is mainly evolutionary in nature. There have been no great revolutions, although there have been important steps, derived from the combination of changes in the environment—instruments,

[21] Cf. Mayr, E., *The Growth of Biological Thought. Diversity, Evolution and Inheritance,* The Belknap Press of Harvard University Press, Cambridge, MA, and London, England, 1982.

research objectives, hybridization of disciplines or fields of knowledge— which lead to the appearance of new disciplines or research lines.

This process is influenced by two important notions which also have an important theoretic weight, evolution and regulation. Its comprehension is difficult by non-experts, which leads to difficulty in the process of socialization of this knowledge through dissemination of mediatic information and even education, due to difficulties experienced by intermediate agents, teachers and journalists, to keep their knowledge updated.

Adjustments between the state of knowledge in biology and socio-political discourses (opinions on economic and ethical aspects) is hindered by the "looping path" followed by advances in this field's knowledge, a path we have pointed out from a combination of historical and philosophical analysis, in line with several previous works in biological philosophy (Mayr, Lewontin).

The new research lines, so-called "omics" —genomics, proteomics— contribute with their data, new counter-arguments with respect to the most extreme reductionist view. These positions enable a distance to be established between two reductionist views: the methodological one providing important contributions to knowledge, and the ideological one which is becoming outdated as the cyclic nature of biological knowledge, going from homogeneity to diversity and viceversa, is revealed.

9. Acknowledgements

This work was performed in the framework of the project entitled "Civil society and the management of science and technology in Spain" (FECYT-CSIC agreement). The previous studies on perception and analysis of contents in mass media received funding from the project under the National R and D Plan BIO 2000-0167-P4-03. The author is a member of the Excellence Network PRIME, funded by the *VI Framework Programme of the European Union*.

10. Bibliography

Dhanda, R. K., "Making the Case for Ethics," *Nature Biotechnology*, v. 22, n. 8, (2004), p. 951.

Diaz Martinez, J. A. and Lopez Pelaez, A., "Biotecnología, periodismo científico y opinión pública: Consenso, disenso, evaluación democrática y difusión de los avances tecnológicos en el siglo XXI," *Sistema,* n. 179-180, March, (2004), pp. 135-158.

Eaton, M. L., *Ethics and the Business of Bioscience*, Stanford University Press, Stanford, 2004.

Echeverria, J., *Los Señores del Aire. Telépolis y el Tercer Entorno,* Ediciones Destino, Barcelona, 1999.

Echeverria, J., *La revolución tecnocientífica,* Fondo de Cultura Económica, Madrid, 2003.

Lewins, R. and Lewontin, R., *The Dialectical Biologist,* Harvard University Press, Cambridge, MA, and London, England, 1990.

Mattick, J. S., "Introns," *Investigación y Ciencia,* n. 339, (December 2004), pp. 26-33.

Mayr, E., *The Growth of Biological Thought. Diversity, Evolution and Inheritance,* The Belknap Press of Harvard University Press, Cambridge, MA, and London, England, 1982.

Muñoz, E., *Biotecnología, Industria y Sociedad: El caso español,* Fundación CEFI, Gabinete de Biotecnología (GABIOTEC), Madrid, 1997.

Muñoz, E., "New Socio-Political Environments and the Dynamics of European Public Research Systems," *Working Paper 02-20*, Science, Technology and Society Group (CSIC), http://www.iesam.csic.es/doctrab2/dt-0220.pdf, 2002.

Muñoz, E., "Percepción pública y Biotecnología. Patrón de conflicto entre información, conocimiento e intereses," in Iañez Pareja, E. (ed.), *Plantas transgénicas: De la Ciencia al Derecho*, Editorial Comares, Granada, 2002, pp. 124 and 129-137.

Muñoz, E., "Ethical Dimensions in Biotech Communications," *EU Focus Workshop on* "*Who should communicate with the public and how?*," Task Group on Public Perceptions, European Federation of Biotechnology, Madrid 13-14 June, 2003.

Muñoz, E., "To Gain Trust: Ethics and Communication in Biotechnology," *Cordia Europabio Convention 2003,* Vienna, 2-4 December, 2003.

Muñoz, E., "Los problemas en el análisis de percepción pública de la Biotecnología: Europa y sus contradicciones," in Rubia Vila, F. J., Fuentes Julian, I. and Casado de Otaola, S. (eds.), *Percepción social de la Ciencia,* Academia Europa de Ciencias y Artes and UNED Ediciones, Madrid, 2004, pp. 159-162.

Muñoz, E., "Opinión pública y Biotecnología: Un 'puzzle' con muchas y variadas piezas," *Sistema,* n. 179-180, March, (2004), pp. 3-13.

Muñoz, E., Plaza, M., Ponce, G., Santos, D. and Todt, O., "Alimentos transgénicos y seguridad. Información y conocimiento en la opinión de los consumidores españoles sobre alimentos transgénicos," *Revista Internacional de Sociología* (RIS), v. 41, May-August, (2005), pp. 93-108.

Muñoz, E. and Plaza, M., "Imágenes de la Ciencia y la Tecnología en España a través del espejo de la encuesta de percepción 2004," *Percepción social de la Ciencia y la Tecnología en España-2005,"* FECYT, Madrid, 2005, pp. 135-161.

Ordoñez, J., Navarro, V. and Sanchez Ron, J. M., *Historia de la Ciencia,* Espasa-Calpe/Austral, Madrid, 2003

Plaza, M., "Análisis de contenido sobre el tratamiento de las aplicaciones biotecnológicas en la prensa española," *Sistema,* n. 179-180, March, (2004), pp. 171-186.

Sanchez Ron, J. M., *El siglo de la Ciencia*, Taurus, Madrid, 2000.

Santos Benito, D. and Dordoni, P., "Opinión pública y debate ético-social sobre un reto de la Biotecnología: Los xenotrasplantes," *Sistema,* n. 179-180, March, (2004), pp. 187-205.

Serres, M. (ed.), *Elements d'histoire des sciences,* Bordas Cultures, Paris (printed in Poitiers), 1989.

The New Encyclopedia Britannica, *The Biological Sciences,* v. 14, Fifteenth Edition, 1990, pp. 920-977.

Turner, L., "The Tyranny of 'Genethics'," *Nature Biotechnology,* v. 21, n. 11, (2003), p. 1282.

Turner, L., "Bioethi$ Inc," *Nature Biotechnology*, v. 22, n. 8, (2004), pp. 947-948.

Cognitive Approach on the Relation Science-Technology[*]

Anna Estany

As the philosophy of science has broadened its horizons and has gone beyond questions of a purely epistemological nature, one of the fields which it has taken an interest in is technology. The philosophy of science has principally dealt with social, political and ethical aspects which has given rise to the so called "Studies of Science, Technology and Society (*STS*)."

STS studies act as an umbrella group under which all kinds of scientific analysis referring to sociopolitical factors in scientific practice may be found. The philosophy of science also comes under this umbrella. Given that the ideas which have predominated in *STS* studies have been of a social constructivist nature, with relativist and irrationalist elements (although not all the work done in *STS* takes this approach), the present paper will put forward the design sciences and praxology as models which allow us to deal with the contextual factors at work in science (social, political, ethical factors, etc.) from a racionalist point of view. This approach has lead us to make useful distinctions between descriptive sciences which describe the world,[1] design sciences which transform the world and technology (machines or instruments). As a result, the best way to deal with the relationship between science, technology and society is through the design sciences[2] and praxology[3] used as theoretical frameworks for interdisciplinary studies of science.

In general, the relationship between science and technology may be dealt with in two ways, that is, either going from science to technology or vice versa. The former approach is that which has been developed most and which has lead to the study of the social consequences of technological development and, in turn, *STS* studies. The latter approach is just as important as the former and has lead to technological innovation being of benefit to scientific development and the development of knowledge in general. We might call this *TIS* (Technological Innovation in Science). In practice, however, these two types of relationship are not independent of each other. There is, in fact, a scientific/technological inter-relationship at work here. The distinction between the two approaches is conceptual in the sense that the study of the consequences of certain technological products for society is not the same as the analysis of the role of technological instruments in scientific research. Nevertheless, technological construction cannot remain indifferent to the cognitive processes of the users of this technology. Most studies carried out on the philosopy of

[*] I am grateful to the Ministry of Education and Science of Spain for sponsoring this work (project BFF2003-09579-C03-01).

[1] Cf. Estany, A. and Casacuberta, D., *¿EUREKA? El trasfondo de un descubrimiento sobre el cáncer y la Genética molecular,* Tusquets, Barcelona, 2003.

[2] Cf. Simon, H. A., *The Science of the Artificial*, The MIT Press, Cambridge, MA, 3rd ed., 1996, and Niiniluoto, I., "The Aim and Structure of Applied Research," *Erkenntnis*, v. 38, (1993), pp. 1-21.

[3] Cf. Kotarbinski, T., "Praxiological Sentences and How they are Proved," in Nagel, E., Suppes, P. and Tarski, A. (eds.), *Logic, Methodology and Philosophy. Proceedings of the 1960 International Congress*, Stanford University Press, Stanford, 1962, pp. 211-223; and Kotarbinski, T., *Praxiology. An Introduction to the Science of Efficient Action*, Pergamon Press, New York, 1965.

technology have ignored this. Neither *STS* studies nor design sciences have taken into account cognitive factors in their models. The aim of this paper is to demonstrate the importance of cogitive factors both for *STS* studies and the design sciences. As the subject is broad in scope, I shall only focus on certain aspects which demonstrate the importance of the cognitve element in the relationship between science and technology.

1. Cognitive Approach in Design Methodology

The aim of the design sciences, in contrast to the descriptive sciences, is to transform the world by means of available scientific knowledge. Examples of design sciences are engineering, architecture, medicine and education. I shall focus on design methodology and praxiology, analyzing how cognitive sciences may be of relevance to the questions raised by these sciences.

The models of design methodology make reference to social, political and economic factors.[4] We shall take McCrory's model as an example of design methodology.[5]

This model has two entry points. The first entry point is the state of the question of basic and applied research relevant to design while the other is the needs generated by non-technical factors (economic, social and geopolitical factors). McCrory also emphasises the differences between the standard scientific method and the design method. The fundamental difference lies in this double entry point which comes from scientific knowledge and human needs. Contextual elements which give rise to human needs do not intervene in the standard scientific method characteristic of pure science or basic research.

As can be seen, no reference is made in McCrory's model to cognitive factors. We might be lead to believe that cognitive factors are included in technical factors. However, there is no indication, either in the model or in other papers related to design methodology, that design takes cognitive factors into account. Neither do the aforementioned authors, Simon and Niiniluoto, who have dealt with the design sciences, make reference to the cognitive element. Indeed, if we look at the whole *STS* field, we shall not find any trace of cognitive factors except where they are conditioned by social factors.

I will argue that it is inadequate to merely include the cognitive in non-technical factors (social, political and geopolitical) as identified by McCrory. Given the two-way direction of the relationship between science and technology, cognitive factors should be considered as a third entry point. Moreover, this third entry point would also have a bearing on the first entry point, that is, the state of the scientific question and on the second entry point, that is, human needs.[6]

The third entry point would therefore be "cognitive processes," which would include emotions, the user's cognitive abilities and psychological factors. Therefore, in phase 2

[4] Cf. Nadler, G., "An Investigation of Design Methodology," *Management Science*, v. 13, n. 10, (1967), pp. B642-B655; Hall, A. D., "Three-dimensional Morphology of Systems Engineering," in Rapp, F. (ed.), *Contributions to a Philosophy of Technology*, Reidel, Dordrecht, 1974, pp. 174-186; Asimov, M., "A Philosophy of Engineering Design," in Rapp, F. (eds.) *Contributions to a Philosophy of Technology*, pp. 150-157; and McCrory, R. J., "The Design Method-A Scientific Approach to Valid Design," in Rapp, F. (ed.), *Contributions to a Philosophy of Technology*, pp. 158-173.

[5] See Figure 1.

[6] See Figure 2.

or the design phase, we need to consider not only the state of the question of scientific knowledge and the needs to be met, but also the cognitive processes of scientists and of the people who will use the designed product and their motivation for doing so.

The design phase, in which a general image of the system is brought into focus, might be conceived as a mental model which is put to the test starting with varied combinations of technological machines. This then leads us to phase 3 in which the viability of the design and the fulfillment of needs, that is, the reason for which the product was created, is confirmed or not. If the design is deemed to be viable we move on to the next phase, phase 4, which is the production and marketing of the product.

In each phase, questions arise in which cognitive factors play an important role. One of these questions relates to the indicators of the product's viability and, above all, ability to fulfill needs. What are the indicators that allow us to conclude that the design fulfills needs? One of them is whether the design makes it easy for the product to be used and it is here that cognitive factors play an important role. Inability to fulfill needs would mean failures in the system. According to McCrory's model, however, such failures are due only to the state of the question. In other words, failure may occur not only in direction of phase 3 to the state of the question but also from phase 3 to the cognitive entry point due to difficulty of use. Finally, in the production and marketing phase, cognitive and emotional factors play a very important role in the acceptance of a product. Here, modes of cognition (experiential and cognitive), as we shall see later on, are very important indicators of success or failure.

2. The Importance of Cognitive Models

McCrory's ideas, as outlined above, were published in 1963. The inclusion of non-technical factors in his model marked an important breakthrough for engineering and, in general, for the applied sciences. Currently, there are many who believe that the cognitive sciences cannot be kept apart from technological design. Once again, engineering is the focus of attention in this new stage of design methodology considered from the cognitive perspective. Therefore, it is hardly surprising that one of the writers who has worked in this field, D. Norman refers to "cognitive engineering"[7] as cognitive science applied to the design and construction of machines and points out the importance of taking into account the fundmental principles that underpin human action with a view to designing systems that require the least possible effort on the part of the user.

One of the first difficulties that cognitive engineering encounters is the difference between psychological and physical variables. People need to control a series of physical systems in order to achieve the aims or objectives that they have set themselves. The person who is going to use the machine must interpret the physical variables in order to achieve his or her aims. To do this, he or she will need to carry out specific actions and handle specific mechanisms.

Any given task, no matter how simple it may be, requires the consideration of various factors and, especially, the gap between a person's aims expressed in psychological terms and the physical system defined in terms of physical variables.

[7] Norman, D. A., "Cognitive Engineering," in Norman, D. A. and Draper, S. W. (eds.), *User Centered System Design. New Perspectives on Human-computer Interaction*, Erlbaum, Hillsdale, NJ, 1986.

Norman puts forward an outline of how the gap between aims and the physical system may be bridged (Figure 3). Each bridge is one-directional, that is, the execution of a task moves from the aims to the physical system while the evaluation of a task moves in the opposite direction. The first bridge begins with the intentions of the user, then moves to the specification of actions and finally contact is made with the physical system. The second bridge involves interpreting the state of the system by comparing it with the original aims and intentions of the user.

Norman applies this outline to the computer and establishes seven phases,[8] which, in practice, may not always appear exactly in the order indicated and at times may omit or repeat some phases. This omission or repetition of phases is due to the complexity of any given task, no matter how simple it may appear. It is important to remember that the gap may be bridged either by bringing the system to the user or the user to the system. Ultimately, the direction taken will depend on whether we wish to prioritize the user or the system.

One way of bridging the gap between the aims and the physical system is by means of a conceptual model of the system which acts as a kind of scaffolding on which to build the bridge. The conceptual model as a mental model provides us with a very clear idea of how designer, user and system interact.[9] As the outline indicates, the designer builds a conceptual model according to the characteristics of the user and gives it material form in a physical system from which image the user builds his or her conceptual model. In other words, the user's conceptual model is the result of the interpretation of the image of the system.

As a result, the designer must try to match the user's model to the design model. In this respect, the designer carries a lot of responsibility given that the user constructs his or her model through the image of the system. The more explicit, intelligible and defined the image is the better able is the user to understand the image of the system. To a large extent, the explicitness of the image depends on the designer although factors such as the type of machine and function it has may also obscure the image. In general, however, the designer must aim to build physical systems that are as visible and intelligible as possible. Norman points out that the image of the system is even more important since many people do not read the instructions although they are incorporated into the system.[10]

Where do we situate these ideas on cognitive engineering in design methodology? Taking McCrory's model, it may be argued that, in the same way that physical and chemical theories (in the case of engineering) depend on the state of the question and sociological and anthropological ideas depend on needs, cognitive processes depend on the principles of design and that this constitutes a new discipline which draws from psychology and computer science.[11]

Given individual and cultural differences, a science of universal design might seem an impossible challenge. Nevertheless, it is worth trying although it may be an "approximate science" in the words of Norman. Designs fail because engineers and designers are too

---

8 See Figure 4.

9 See Figure 5.

10 Cf. Norman, D. A., "Cognitive Engineering," in Norman, D. A. and Draper, S. W. (eds.), *User Centered System Design. New Perspectives on Human-computer Interaction*, p. 47.

11 Cf. Norman, D. A., "Cognitive Engineering," p. 59.

centred on themselves. The former tend to focus on design in technology. In fact, they know too much about technology and too little about the needs of ordinary people. It is far more important to observe how people behave, how they fail than to carry out surveys in order to determine what exactly people need.[12] Norman here refers to ordinary people, however, we could adopt the same approach with scientists by observing them in their laboratories, at conferences, in class in order to determine their likes, abilities, etc.

All of the above examples and models are conceived with engineers in mind and, especially computer engineering in the case of Norman's model. Nevertheless, any design principle may be applied to other fields. Indeed, the model that relates designer, user and system could be anything from a tap, plugs, the dashboard of a car, a list of medicine to be taken to a whole range of products that come from the application of scientific knowledge to the satisfaction of human needs.

3. A Cognitive Praxiology?

We have seen how design methodology would be affected by the incorporation of cognitive factors. We have established that the theoretical basis of design is praxiology, that is, the science of efficient action. Now we shall try to determine where cognitive sciences might make important contributions to the three elements which Kotarbinski considers to be fundamental to praxiology, that is, the theoretical, technical and behavioural base. Praxiological statements are linguistic forms with which praxiology expresses itself. A praxiological statement takes the following form:

If you want to reach A and you are in situation B you must do C

If you want to lose weight and you eat a lot of meat, cakes, etc., you should reduce fat consumption by half and replace it with fruit and vegetables

The theoretical base is that which establishes the fact that C causes A being in situation B (if you reduce consumption by half and also eat more vegetables you will lose weight). We would also need to include here biological and chemical knowledge which is the theoretical base of dietetics (a design science). Here the importance of cognitive factors derives from *TIS*. In other words, technological innovation with regard to instruments and techniques in scientific research can bring about development in design sciences which, in turn, can make it possible to fulfill the proposed objectives of the design science in question in a more efficient way. In the above case of dietetics, this would be a more suitable diet that would allow one to lose weight without negative side effects on one's physical and mental health.

The technical base refers to all the instruments and techniques necessary in order to achieve the objective. In the example of the diet, it would appear that there aren't any techniques in the strict sense of the word other than operations to reduce the size of the stomach or which use machines to carry out liposuctions, etc. However, the technical base should be understood in broader terms and, as a result, we may consider any diet plan as a technique to lose weight. From the cognitive perspective, Norman's outline of the relationship between designer, user and system is valid in this context. A technique may be easily handled or understood by the user (what is referred to as "patently obvious") or may require the skill of an expert (as in the case of the machine) or careful attention (as in the case of the diet plan).

[12] Cf. Norman, D. A., *Emotional Design. Why We Love (or Hate) Everyday Things*, Basic Books, New York, 2004.

The behavioural base refers to the actions required to achieve the objective. This involves the creation of a hierarchy of objectives which can go from the long term objective of improving one's physical appearance to the immediate objective of eating an apple and drinking a coffee for breakfast passing through the central and short term objective of losing weight. As a result, what is important here is a theory of action in order to achieve proposed objectives and devise a plan which will lead to the corresponding actions. From the cognitive perspective, Miller's, Galanter's and Pribram's model of the relationship between plans and behaviour is important here.[13] The fundamental idea of the model is the "TOTE unit" (Test-Operate-Exist-Unit).[14]

Energy, information and control flow through the arrows. In neurological terms, these arrows are the neuronal impulses which are produced when information coming from the environment reaches the brain through the senses and controls the organism.[15] "The TOTE unit" simply confirms that the operations carried out by an organism are constantly guided by the results of various tests.[16] When actions are more complex we have units within a more general outline in such a way that we end up with a hierarchy of units which correspond to a hierarchy of plans which is how the organization of behaviour might be described.

Although the above authors do not refer to mental models the plans which guide our actions are models for action. The TOTE model may be of relevance to Norman's cognitive models on the relationship between the designer, user and physical system and on the bridges between aims and physical system. In both cases, it is a question of plans that are subordinate to other plans and, in the end, what we end up with is a hierarchy of objectives which involve a series of actions. Thus, Miller, Galanter and Pribram provide us with a model from the perspective of cognitive psychology underlies these cognitive engineering models which, in turn, underlies the whole design process.

4. Relationship Between the Social and the Cognitive

The introduction of the cognitive element should not be seen as something that is independent of and unconnected to the contextual elements (social, political, cultural, ethical, etc.) which STS studies has focussed on (not to say limited itself to). One of the situations in which the relationship between the cognitive and the social is clearly manifest is when errors are made in the course of any technological process, which bring about personal and social misfortune and which, in many cases, are due to "human errors." As the saying says "to err is human," however, it is also human to try and avoid error. In general, error is understood as a failure to achieve the desired result based on a planned sequence of physical and mental activities. These errors may be due to slips or mistakes as a result of some failure in a series of actions or due to defects in deductions which are made between decisions taken and the desired objective. There is no doubt that the latter errors are more complex and serious than the former and that they may have more serious consequences.[17]

---

[13] Cf. Miller, G. A., Galanter, E. and Pribram, K., *Plans and Structure of Behavior*, Holt, Rinehart and Winston, New York, 1960.

[14] See Figure 6.

[15] Cf. Estany, A., *Vida, muerte y resurrección de la conciencia. Análisis filosófico de las revoluciones científicas en la Psicología contemporánea*, Paidós, Barcelona, 1999, p. 169.

[16] Cf. Miller, G. A., Galanter, E. and Pribram, K., *Plans and Structure of Behavior*, p. 29.

[17] Cf. Reason, J., *Human Error*, Cambridge University Press, Cambridge, 1990.

How may errors be avoided? Design (good design) is one way, amongst others. The perfect design does not exist, however, "the good design" does and in order to achieve this, a series of requisites need to be met.[18] Petroski points out that the analysis of documented cases might be useful to engineers given that failures and successes might be used to construct a theoretical framework.[19] Moreover, these cases might help to make engineering students aware of the roots of their profession and its relationship to society. In short, they might learn from the experience of previous generations.

The concept of failure is central to the design process and only by avoiding failure do we achieve successful designs. There is a saying in engineering and design circles which says that more is learnt from failures than from successes. According to Lev Zetlin, engineers should be paranoid during the design process. One engineer commented: "I expect disaster, I'm afraid that a tragedy will occur." The engineers, R. Stephenson and H. Hoover would say: "we cannot sleep at night because we take the design problems to bed with us."[20]

Errors made in the course of any human activity is just one example of how the cognitive and social converge. As a result, criticism of the *STS* and *Design Science* models due to the fact that they limit themselves to sociopolitical factors does not mean that we should ignore them rather we should try to incorporate them into a new theoretical framework. All of these ideas can be more easily demonstrated and studied in the engineering sciences. However, it is possible to extend these ideas to other areas such as the operation of an X-ray machine in a hospital, the design of traffic signs or a television remote control. A bad design may cause errors with more or less serious consequences, however, in either case, there is a negative social effect.

5. Cognitive Perspective of Technological Innovation in Science (*TIS*)

TIS is understood as a process through which technological innovation benefits science itself and, consequently, knowledge. As pointed out at the beginning of this paper, this is another way of understanding the relationship between science and technology. The relationship is not usually understood in this way due to the fact that the origin of this relationship is always situated in science which produces technology and which has social consequences. Instead of reaching its end point in society, technological innovation benefits science itself thus producing feedback. Pure science is the foundation of design sciences, which scientists use (positively speaking) to obtain finance for basic research. Let us not forget that no matter where the financial support comes from, whether it be private or public in origin, it will be easier to obtain if the reseacher links it to a design science. For example, basic research in molecular biology has undoubtedly benefited from the fact that it is the theoretical base of cancer research. Pure science makes it possible to construct technology, part of which will be used as scientific instruments in research carried out in pure science itself. Thus, the circle is closed.

---

[18] Cf. Petroski, H., *Small Things Considered. Why There is no Perfect Design*, Alfred A. Knopf, New York, 2003.

[19] Cf. Petroski, H., *Design Paradigms. Case Histories of Error and Judgment in Engineering*, Cambridge University Press, Cambridge, MA, 1995.

[20] Quoted by Petroski, H., *Design Paradigms. Case Histories of Error and Judgment in Engineering*, p. 3.

The above idea refers to the role played by scientific instruments in scientific research. Nobody has denied the importance of instruments (neither scientists, historians nor philosophers), however, the degree of importance given to them varies. What perhaps is new here is that the cognitive element plays a part in the ITS model in that the success of instruments or machines is based on their construction according to design principles put forward by Norman.

Although the focus of this paper is scientific research the *TIS* model also refers to knowledge. As a result, contributions made by cognitive sciences as to the characteristics that technological machines should have so that they match the cognitive abilities of the user refer to theoretical knowledge in general and not scientific knowledge in particular. Moreover, taking naturalist philosophy as a standpoint, even in its minimlist form,[21] it follows that what is relevant to knowledge in general is also relevant to scientific knowledge and that, ultimately, scientists are human beings and thus subject to the same cognitive limitations as all other human beings. The only difference between scientists and other human beings is that scientists are more predisposed to learning and being experts in specific subjects, that is, those subjects that are relevant to their research. This means that the experience and skills accumulated in their training as scientists may alter the image of the system that the designer atrributes to the scientist/user. It does not mean that the designer should not try to make the design model compatible with the scientist/user model and that in order to achieve this he or she make the image of the system explicit, transparent and intelligible. The difference between them is that what is intelligible and transparent to one is not necessarily so to the other. These differences are true both amongst scientists working in different fields and amongst scientists and non-scientists.

What makes *TIS* especially interesting is the fact that we are immersed in a technologized world, both at the social and personal level. The idea of "cyborg," that is, a simbiosis of flesh and machine, as attributed by Andy Clark to the thinking units,[22] is an example of the personal implications of technology. I shall not analyze this phenomena from the perspective of the philosophy of mind or philosophical anthropology although it undoubtedly has important repercussions here. Nevertheless, it is important to point out that any machine which is an extension of our mind contributes, either positively or negatively, to the acquisition of knowledge.

The crucial point here is that we are constantly interacting with machines and that all questions related to technological design concern us as human beings and as a result, what distinguishes us from other animals is, intelligence. In this way, technological machines, ranging from a pencil to a calculator, may be seen as an extension of our cognitive abilities.

Both Norman and other authors who have dealt with interaction between cognition and technology have centred their attention on the computer. Indeed, one of Norman's and Draper's work focusses on the design of computers and demonstrates that the computer,[23]

[21] Cf. Estany, A., "The Theory-laden Thesis of Observation in the Light of Cognitive Psychology," *Philosophy of Science*, v. 68, (2001), pp. 203-217.

[22] He develops this idea in Clark, A., *Natural-born Cyborgs. Minds, Technologies, and the Future of Human Intelligence*, Oxford University Press, Oxford, 2003.

[23] Cf. Norman, D. A. and Draper, S. W. (eds.), *User Centered System Design. New Perspectives on Human-computer Interaction*, passim.

in the same way as any other design, must take into account technology, the person and interaction between the two. Many works have been published in recent decades on the adaptation of the physical surface of the computer to the skills of the user. Given the fact that computer programmes are used on a large scale in all fields of scientific research we should aim to make the computer more accessible to the user. This greater accessibility, in turn, would undoubtedly benefit the development of science. In fact Norman exemplifies the relationship between the designer and any physical system.

Most cognitive models have elements which are relevant to the acquisition of knowledge and scientific research. I shall take Norman's contributions, which I believe are especially relevant to *TIS*, in order to demonstrate the relationship between cognition and technology.

5.1. Modes of Cognition

Norman distinguishes between two modes of cognition: experiential and reflective.[24] The experiential mode allows us to perceive and react to events in an efficient and effortless way whereas the reflective mode allows us to compare and contrast, think and make decisions. Nevertheless, we should not see these two modes as being independent of each other but rather as two elements in our lives as thinking beings which we need and use. Technological instruments may help us or make our life difficult depending on the appropriateness of the design. Many problems come from the fact that machines which have been designed for experiential behaviour require thought. As a result, simple tasks become difficult as they require thought and effort. For example, a camera requires us to act quickly, however, if we need to think too much the result will be failure, that is, we won't achieve the objective for which we bought the camera. Similarly, we may encounter problems with machines that require thought but do not allow us to compare and make decisions. This is the case of electronic machines that restrict the availability of information which makes it difficult to make comparisons and thus to make decisions. The fact that printed copies are handed out in many meetings despite the fact that everybody has access to a personal computer is another example of this difficulty. In order to tackle this problem, we should be able to have various screens open at the same time.

5.2. Levels in the Brain

Another aspect of cognition is the levels of the brain, the result of human attributes,[25] and which are related to modes of cognition. The visceral level is automatic and genetically programmed. At this level, we are all similar although there are differences, for example, we all have a certain fear of heights, however, for some it would be impossible to look through the rungs of a ladder while others are able to climb mountains. The behavioural level is the part that contains brain processes which control everyday behaviour. At this level, there are many more differences between us and we are more influenced by experience, training and education. The reflective level is the contemplative part of the brain. Culture plays a

[24] Cf. Norman, D. A., *Things that Make Us Smart. Deffending Human Attributes in the Age of the Machine*, Addison-Wesley, Reading, MA, 1993.

[25] Cf. Norman, D. A., *Emotional Design. Why We Love (or Hate) Everyday Things*, p. 21.

very important role at this level although there are some universal characteristics such as the fact that most adolescents tend to dislike what adults like.

These three levels should be taken into account when designing technology. Norman applies these three levels to distinct characteristics of products.[26] The visceral design takes into account appearance and that which produces positive vibrations in us according to our nature as human beings. The behavioural design is aimed at achieving effective use and thus the first test for this type of design is whether it satisfies the needs for which it has been designed. The reflective design is aimed at self-image, personal satisfaction and memories. As a result, this design takes into account cultural factors. It is not practical or biological rather everything is in the mind of the person who possesses the product.

Norman puts forward these ideas within the context of emotions and the role they play in design theory, which is not especially conceived with scientific instruments in mind. Nevertheless, we may argue that modes of cognition, the levels of the brain and emotions are relevant to scientific research.

6. Repercussions for science

Firstly, we may ask ourselves whether the handling of scientific instruments requires an experiential or reflective mode of cognition. On the one hand, everything would seem to indicate that it requires a reflective mode since scientific research involves making decisions and comparisons. However, let us not forget that for science instruments are precisely instruments that allow us to achieve an objective, that is, the explanation of phenomena, in a precise, quick and productive way. As a result, technological instruments or machines should be easy to use and are not an end in themselves.

Of the three levels of the brain, the behavioural level would seem to be the one which is most in tune with effective use. This does not mean, however, that the other levels are not important. Scientists are also more favourably disposed to certain colours, shapes, etc. In other words, they have aesthetic preferences and, of course, they are immersed in a culture which has given them different experiences and abilities. The above considerations are important when it comes to using a machine. A scientist would probably not use the same criteria when buying a machine for the laboratory as when buying a piece of furniture for the house. In other words, it would be better to keep the visceral and cultural levels in the background although they should not be forgotten.

Empirical studies on the relationship between cognition and technology have until now had very little to do with scientific research. The laboratory has not been studied as have the cabine of planes and the engine room of boats.[27] The laboratory has been studied by the social constructivists in order to analyze relationships and social structures,[28] and, in the majority of cases, to demonstrate the lack of objectivity in science.

[26] Cf. Norman, D. A., *Emotional Design. Why We Love (or Hate) Everyday Things*, pp. 36-38.

[27] Cf. Hutchins, E., *Cognition in the Wild*, The MIT Press, Cambridge, MA, 1995.

[28] Cf. Knorr-Cetina, K. D., *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*, Pergamon, Oxford, 1981; and Latour, B. and Woolgar, S., *Laboratory Life: The Social Construction of Scientific Facts*, Sage, London, 1979.

7. Conclusions

- The philosophy of technology should include the cognitive element in its models.
- Design methodology should take into account what cognitive psychology says about mechanisms for processing information.
- One line of research within the design sciences should be to investigate the causes of errors made by people who operate machines.
- The above conclusions lead us to conclude that cognitive sciences will play a very important role in design sciences since human needs intervene in all of them. Human beings have aims and intentions and act in order to satisfy their needs.
- *TIS* points us to the relevance which the above has for scientific research itself. This debate emerges when the philosophy of science goes beyond epistemological values and takes an interest in contextual factors but ignores the cognitive aspect. We have yet to see how technological design will return to the epistemological element. Given that instruments and technology have played a very important role in scientific research, everything that is related to design technology will have repercussions on the research that uses this technology.

8. Bibliography

Asimov, M., "A Philosophy of Engineering Design," in Rapp, F. (eds.), *Contributions to a Philosophy of Technology*, Reidel, Dordrecht, 1974, pp. 150-157.

Clark, A., *Natural-born Cyborgs. Minds, Technologies, and the Future of Human Intelligence*, Oxford University Press, Oxford, 2003.

Clark, A. and Chalmers, D., "The Extended Mind," *Analysis*, v. 58. n. 1, (1998), pp. 7-19.

Estany, A., *Vida, muerte y resurrección de la conciencia. Análisis filosófico de las revoluciones científicas en la Psicología contemporánea*, Paidós, Barcelona, 1999.

Estany, A., "The Theory-laden Thesis of Observation in the Light of Cognitive Psychology," *Philosophy of Science*, v. 68, (2001), pp. 203-217.

Estany, A., and Casacuberta, D., *¿EUREKA? El trasfondo de un descubrimiento sobre el cáncer y la Genética molecular,* Tusquets, Barcelona, 2003.

Fauconnier, G. and Turner, M., *The Way We Think. Conceptual Blending and the Mind's Hidden Complexities*, Basic Books, Oxford, 2002.

Hall, A. D., "Thre-dimensional Morphology of Systems *Engineering*," in Rapp, F. (ed.), *Contributions to a Philosophy of Technology*, Reidel, Dordrecht, 1974, pp. 174-186.

Hutchins, E., *Cognition in the wild*, The MIT Press, Cambridge, MA, 1995.

Jacques, R. and Powell, J. A. (eds.), *Design: Science: Method*, Westbury House, Guilford, Surrey, 1981.

Knorr-Cetina, K. D., *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*, Pergamon, Oxford, 1981.

Kotarbinski, T., "Praxiological Sentences and How they are Proved," in Nagel, E., Suppes, P. and Tarski, A. (eds.), *Logic, Methodology and Philosophy. Proceedings of the 1960 International Congress*, Stanford University Press, Stanford, 1962, pp. 211-223.

KOTARBINSKI, T., *Praxiology. An Introduction to the Science of Efficient Action*, Pergamon Press, New York, 1965.

LATOUR, B. and WOOLGAR, S., *Laboratory Life: The Social Construction of Scientific Facts*, Sage, London, 1979.

MCCRORY, R. J., "The Design Method-A Scientific Approach to Valid Design," in RAPP, F. (ed.), *Contributions to a Philosophy of Technology*, Reidel, Dordrecht, 1974, pp. 158-173.

MILLER, G. A., GALANTER, E. and PRIBAM, K. H., *Plans and the Structure of Behavior*, Holt, Rinehart and Winston, New York, 1960.

NADLER, G., "An Investigation of Design Methodology," *Management Science*, v. 13, n. 10, (1967), pp. B642-B655.

NIINILUOTO, I., "The Aim and Structure of Applied Research," *Erkenntnis*, v. 38, (1993), pp. 1-21.

NORMAN, D. A., *Learning and memory*, W. H. Freeman and Company, New York and Oxford, 1982. Spanish translation by María Victoria Sebastián Gascón and Tomás del Amo: *El aprendizaje y la memoria*, Alianza Editorial, Madrid, 1985.

NORMAN, D. A., "Cognitive Engineering," in NORMAN, D. A. and DRAPER, S. W. (eds.), *User Centered System Design. New Perspectives on Human-computer Interaction*, Erlbaum, Hillsdale, NJ, 1986, pp. 31-61.

NORMAN, D. A., *The Psychology of Everyday Things*, Basic Books, New York, 1988. Spanish translation by F. Santos Fontenla: *La Psicología de los objetos cotidianos*, Editorial Nerea, Madrid, 1990.

NORMAN, D. A., *Turn Signals are the Facial Expressions of Automobiles*, Addison-Wesley, Reading, MA, 1992.

NORMAN, D. A., *Things that Make Us Smart. Defending Human Attributes in the Age of the Machine*, Addison-Wesley, Reading, MA, 1993.

NORMAN, D. A., *Emotional Design. Why We Love (or Hate) Everyday Things*, Basic Books, New York, 2004.

PETROSKI, H., *To Engineer is Human. The Role of Failure in Successful Design*, St. Martin's Press, New York, 1982.

PETROSKI, H., *Design Paradigms. Case Histories of Error and Judgment in Engineering*, Cambridge University Press, Cambridge, MA, 1995.

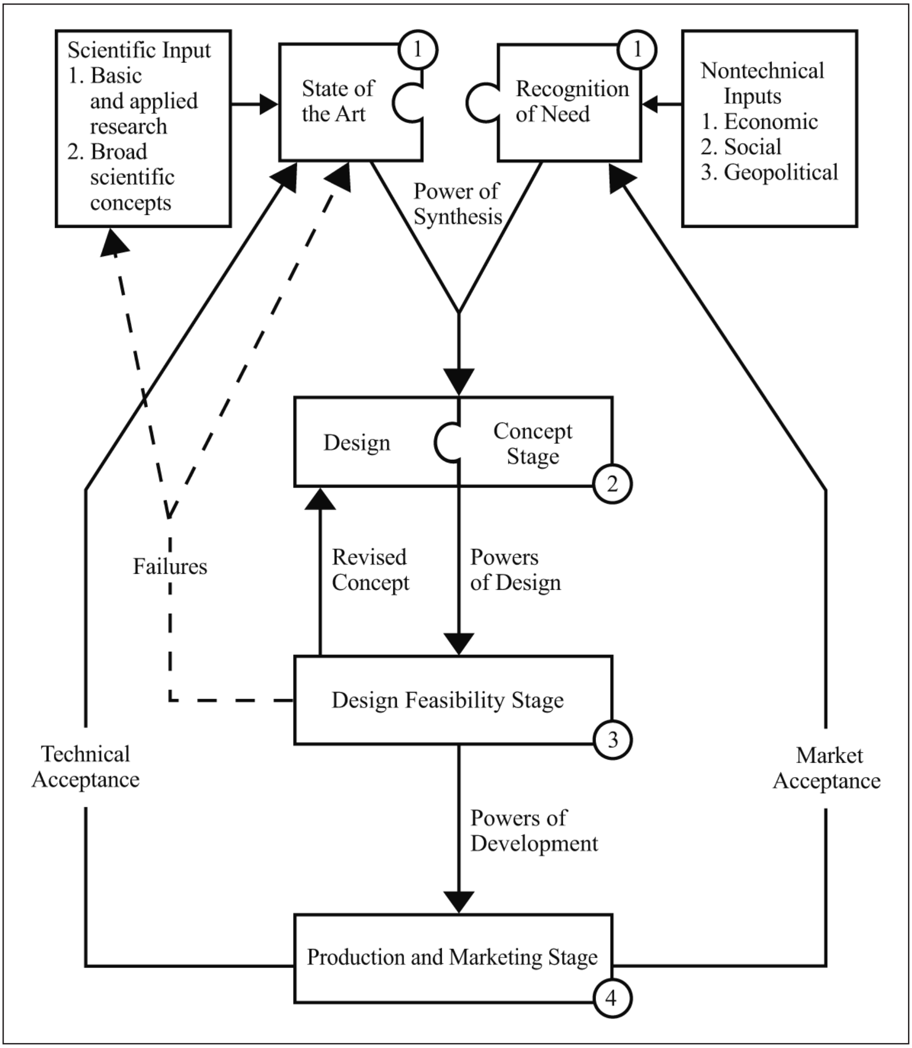
PETROSKI, H., *Small Things Considered. Why There is no Perfect Design*, Alfred A. Knopf, New York, 2003.

REASON, J., *Human Error*, Cambridge University Press, Cambridge, 1990.

SIMON, H. A., *The Sciences of the Artificial*, The MIT Press, Cambridge, MA, 3rd ed., 1996.

FIGURES

Figure 1. Graphic representation of the design method.



Cf. McCrory, R. J., "The Design Method-A Scientific Approach to Valid Design," in Rapp, F. (ed.), *Contributions to a Philosophy of Technology,* p. 162.

Figure 2. Reformulation of the design method.

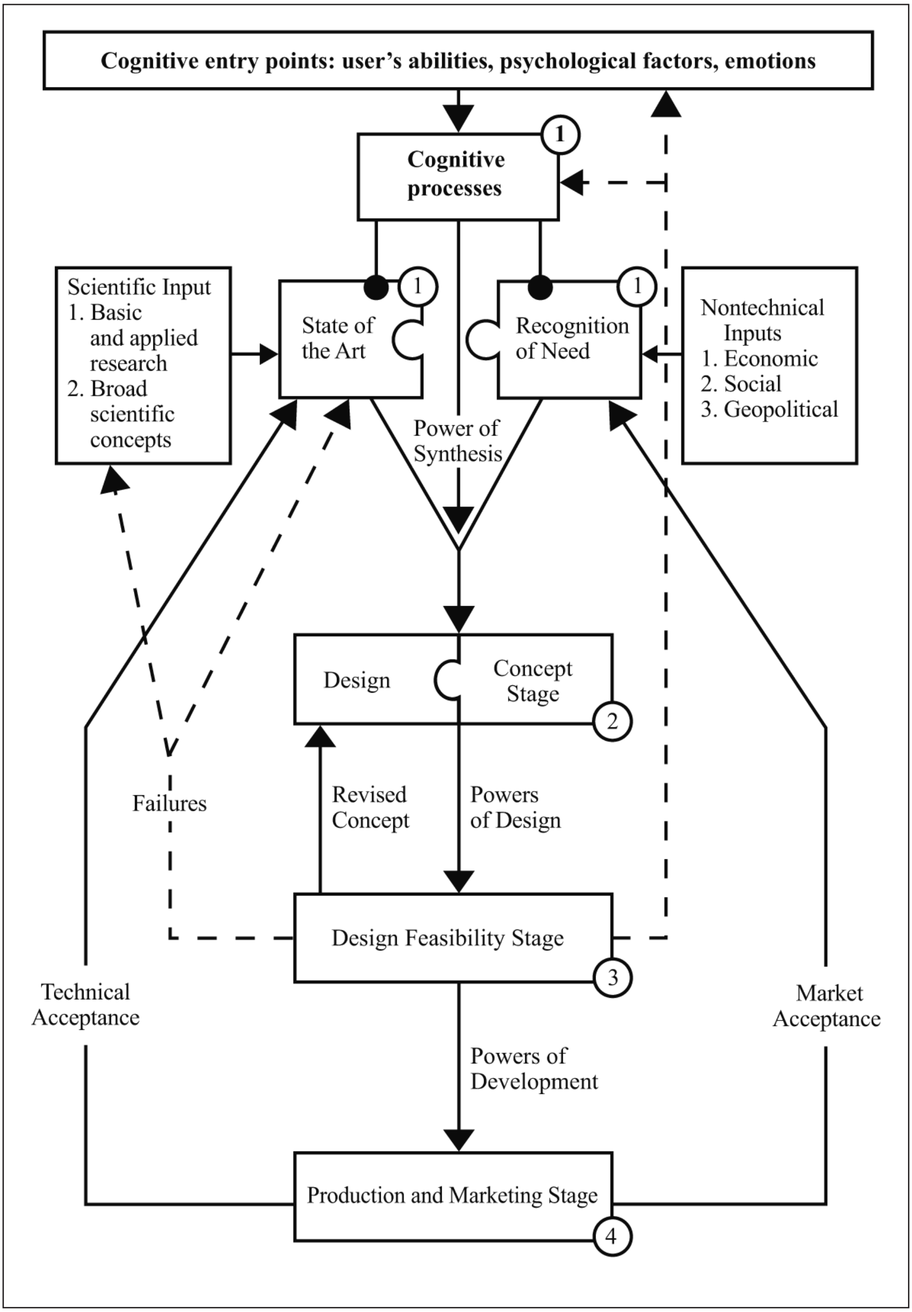
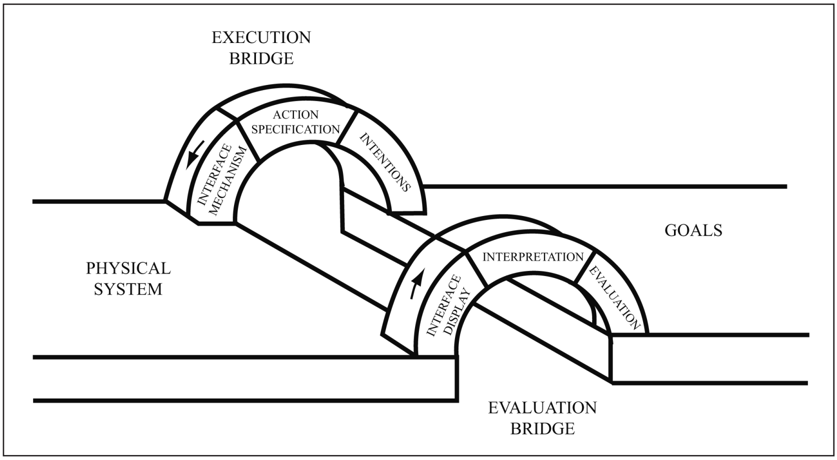
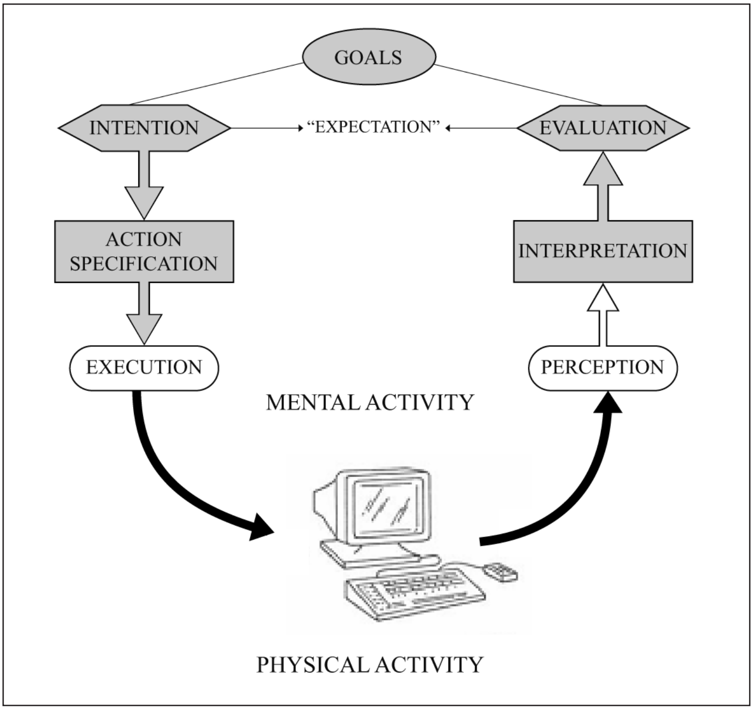


Figure 3. Execution bridge and evaluation bridge
which bridge the gap between goals and physical system.



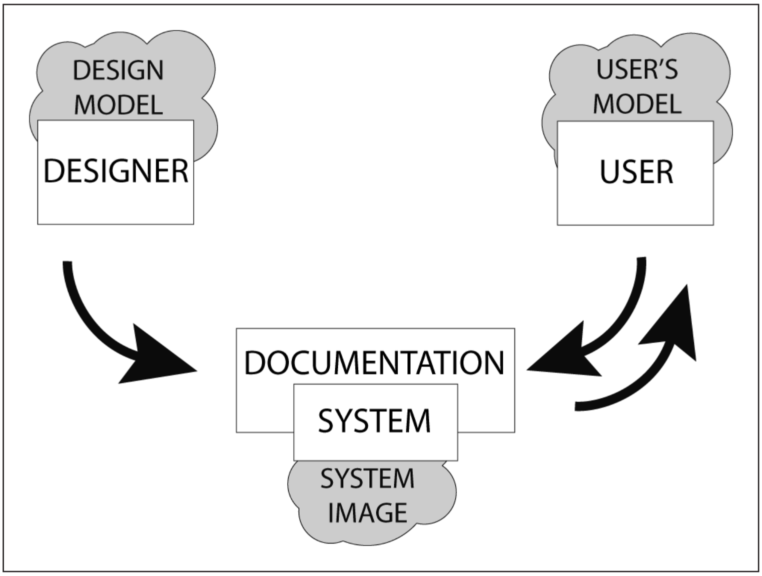
Cf. NORMAN, D. A., "Cognitive Engineering," p. 40.

Figure 4. Series of actions taken to reach an objective.



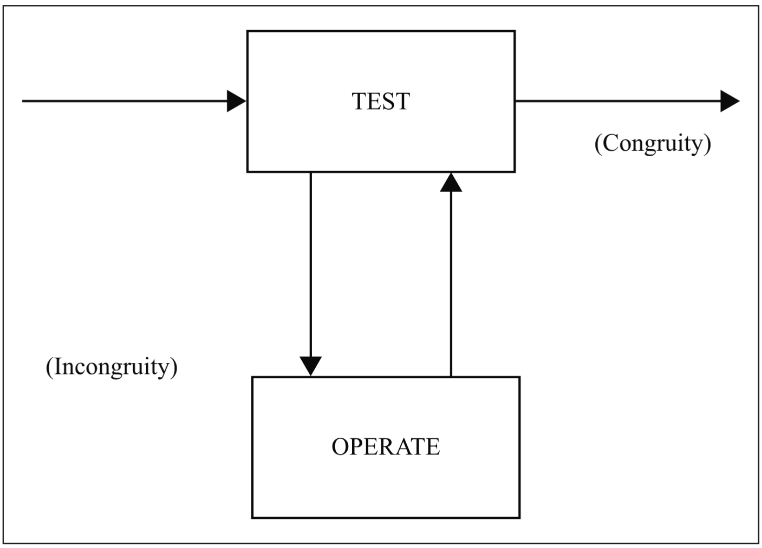
Cf. NORMAN, D. A., "Cognitive Engineering," p. 42.

Figure 5. Relationship between designer, user and physical system.



Cf. NORMAN, D. A., "Cognitive Engineering," p. 46.

Figure 6. Model of unit TOTE (Test-Operate-Test-Exist-Unit).
Congruity, Incongruity, Operate.



Cf. MILLER, G. A., GALANTER, E. and PRIBAM, K. H., *Plans and the Structure of Behavior,* p. 26.

III

Focal Philosophical Problems in Empirical and Formal Sciences

Philosophy and Neuroscience: The Problems

Peter Machamer

1. Philosophy and Neuroscience: The Issues

If one does not hold the view that philosophy is a totally *sui generis* enterprise and if one does not hold that the domain of philosophical inquiry is somehow limited to the *apriori*, abstract objects or some other collection of "phenomena" that is hermetically sealed off from the domains treated by the sciences, then it ought to be that studies in neuroscience have relevance to issues of mind/brain, the nature of knowledge and knowledge acquisition and use, and more general issues in epistemology and metaphysics.

It is also the case that neuroscience, in whatever sub discipline, i.e., cognitive neuroscience, neurobiology, etc., since it is a science, should raise all the problems about explanation, evidence, role of experiment, discovery of mechanisms, etc. that all sciences have. That is, the basic issues in philosophy of science may be raised specifically about neuroscience.

If the above is true, then this means philosophy of neuroscience has two types of problems:

(1) The old philosophical kinds: subjectivity, consciousness, experience, belief and knowledge; and

(2) Problems arising from treating neuroscience as a branch of science, the traditional philosophy of science questions.

There are methodological schools of psychology that hold, just like the behaviorists of old, that psychology need not attend to the "hardware" or "wetware" of what goes in the brain. Any epistemic and cognitive theories may be advanced and tested independently of any neuro-data. Sometimes a computer analogy, specifically the distinction between program level descriptions and logic circuits or even hardware, has been invoked (by Fodor and Dennett) to support such claims.[1] A more empiricist, old positivistic, yet recently contemporary, view holds that all the theory that one needs or is entitled to have will come from analyzing data in the proper way. Using Bayes nets is a favorite here, though other forms of computational modeling subscribe to the similar principles and goals. I shall deal with this radical empiricism only in passing.

Let us first consider the philosophy of science problems. First, what is the nature of neuro-scientific explanations? I think it is correct to claim that most neuroscientists seek to explain the phenomena they are interested in by discovering and elaborating mechanisms,[2] and further that this is how they characterize their own work. However, before we delve into this problem certain preliminary distinctions need to be made clear.

---

[1] Cf. Fodor, J., "Special Sciences (Or: The Disunity Of Science As A Working Hypothesis)," *Synthese*, v. 28, (1974), pp. 97-115, and Dennett, D., *The Intentional Stance*, The MIT Press, Cambridge, MA, 1987.

[2] Cf. Machamer, P. K., Darden, L. and Craver, C. F., "Thinking About Mechanisms," *Philosophy of Science* v. 67, n. 1, (2000), pp. 1-25.

In discussing the sciences of the brain, it will prove convenient to distinguish among neurobiology, cognitive neuroscience and computational neuroscientists. We could also add the medical field of neurology, and number of other research areas.

Philosophically, the most important distinction is between phenomena and data.[3] *Phenomena* are what are being explained, conceived or observed, often in a "pre-scientific" way. "Pre-scientific" means we have somehow isolated a type of event that we deem to be important, and which we then will try to scientifically investigate it through experiment, observation and other means. For example, one wants to explain human memory loss during old age, or what kinds of things are remembered the longest, or how perceptual recognition occurs, or word learning, etc. Biologically, one may wish to find, e.g., out under what conditions and what is it in neurons that changes such that they fire more frequently to types of stimuli.

Within the field of neuroscience it is common to distinguish three areas of research: cognitive neuroscience, neurobiology and computational neuroscience. *Cognitive neuroscience* attempts to discover the neural or brain mechanisms that cause (most often they say, are implicated in) behavior and actions and cognitive functions. Often times use imaging devices (fMRI, nMRI or PET scans) combined with operationalized experimental tasks to attempt to localize of where in the brain these cognitive functions are occurring. Sometimes this area is called neuropsychology. Occasionally a daring researcher will purport to give us information about consciousness, subjectivity, or the self.

Neurobiology investigates the electro-chemical processes that occur at the cellular or neuronal level. Examples are inducing LTP (long term potentiation), examining patterns of dendritic growth, causing neuronal learning (firing to a stimulus type), and finding examples of neuronal plasticity where one set of neurons takes on a function that before had been carried on by another set or another system.

Data, as opposed to the phenomena, are what is the outcome of experimental or measuring procedures used by the scientists during their research. It is most often presumed that the data collected bears an obvious and a straightforward relationship to the phenomena being investigated. This is not always the case. This relation between data and phenomena is most often talked about as the problem of validity (as opposed to reliability.) We may have quite reliable experimental procedures that are inter-subjectively useable, and that yield data that accurately reflect the outcomes of the experiments giving us a coherent self consistent data set, yet they may have no validity *vis a vis* the phenomena of interest. They are measurements, but do not measure what we are really interested in.

In neurobiology the types of data collected usually are measures of the frequency of some behavior or output deemed relevant to a cell or neuron's functioning or resulting from the operation of some mechanism. Very often they are measures of the frequencies of cell and neuronal firing discharges, in the presence of some controlled stimulus or input.

The major experimental paradigms from which data are collected in cognitive neuroscience are repeated stimuli, classical and operant conditioning, priming, and recall and recollection experiments. The theoretical assumptions presupposed by most experimental paradigms are that repeated stimulation and/or classical, or operant conditioning are the

[3] Cf. Bogen, J. and Woodward, J., "Saving the Phenomena," *The Philosophical Review*, v. 97, (1988), pp. 303-352.

operative mechanisms for all learning and memory phenomena. This is questionable belief, given that the cognitive revolution of the 60s challenged the adequacy of exactly these paradigms as being sufficient to account for human cognition.

2. Systems and mechanisms

Most scientists explain most phenomena by discovering, elaborating, and studying mechanisms. In an earlier paper Lindley Darden, Carl Craver and myself put forward a tentative characterizing definition.[4] We said:

"Mechanisms are entities and activities organized such that they are productive of regular changes[5] from start or set-up to finish or termination conditions… Mechanisms are composed of both *entities* (with their properties) and *activities*. Activities are the producers of change… Entities are the things that engage in activities."[6]

Some years after, there is one error in this that needs to be changed. Delete the word "regular" is too law-like, and it is used to make a claim that is just false. There may be, and most certainly are, mechanisms that operate only once. And since we will not allow counterfactuals as part of the analysis, we cannot hold that regularity is necessary. We do not allow counterfactuals because their truth conditions are not clear or non-existent. By the way, this is not the case for many conditional (if…then) statements, and we do allow conditional reasoning (who wouldn't?). But conditionals cannot be part of the analysis of how the mechanism works

Carl Craver wrote in recent draft: "The boundaries of what counts as a mechanism (what is inside and what is outside) are fixed by reference to the phenomenon to be explained."[7] But "fixed" is most often too strong. There are many ways to skin a cat; and, the same protein may be made in so many different ways that it often becomes impossible to form a generalization (or law) that there is a single "gene" responsible or the mechanism used is a particular case is always the mechanism for making that protein.[8] In most any case in neuroscience, the termination conditions allow for many alternative paths that would bring them into being. (Carl Hempel recognized this when talking about functional explanations when he noted that they lack necessity because of alternative mechanisms for achieving the same end.)[9]

A realistic example will help specify this, and further allow us to bring out some more points. Patricia Churchland and Terrance Senjowski report some work that was done on an

---

[4] Cf. Machamer, P. K., Darden, L. and Craver, C. F., "Thinking About Mechanisms," pp. 1-25.

[5] I think "regular" should be dropped from the definition. Jim Bogen has argued forcefully that there might be mechanisms that operated only once and a while or even one that worked only once.

[6] Machamer, P. K., Darden, L. and Craver, C. F., "Thinking About Mechanisms," p. 3.

[7] Craver, C. F., *Explaining the Brain: What a Science of the Mind-Brain Could Be*, Oxford University Press, Oxford, forthcoming. He cites Bechtel, W. and Richardson, R., *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Princeton University Press, Princeton, 1992, as well as Glennan, S., "Mechanisms and the Nature of Causation," *Erkenntnis*, v. 44, (1996), pp. 49-71.

[8] Cf. Griffiths, P. and Stotz, K., "Representing Genes," www.pitt.edu/~kstotz/genes/genes.html, and more generally Mitchell, S. D., *Biological Complexity and Integrative Pluralism* (*Cambridge Studies in Philosophy and Biology*), Cambridge University Press, Cambridge, 2003.

[9] Cf. Hempel, C. G., "The Logic of Functional Analysis," in Gross, Ll. (ed.), *Symposium on Sociological Theory*, P. Row, Evanston, IL, 1959, pp. 271-307; reprinted in Hempel, C. G., *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*, Free Press, N. York, 1965, pp. 297-330.

avoidance response of the leech.[10] The leech is a model organism, they say, because of the simplicity of its central nervous system, the ease of experimenting on it, and the availability of the organism —an interesting mix of kinds of criteria. But our focus is this: there is basically a straightforward decomposition strategy used is examining how the leech performs this aversive reaction of bending when prodded. We break large goal oriented tasks, into sub-tasks performed by subsystems, and the whole is then the mechanism of the overall system.[11] This strategy and the assumptions about the leech's functioning are teleological at every stage. The teleological tone is set by the very first description of the case; "The local bending reflex is the behavior of the leech… is a very simple escape behavior wherein potential harm is recognized and avoidance action taken. Basically the leech withdraws from an irritating stimulus by making a kind of large kink in the length of its hoselike body."[12]

The leech, they report, is a "segmented animal, and within each segment a ganglion of neurons coordinates behaviors.[13] Nine types of interneurons have been identified interposed between the sensory and motor neurons, which mediate dorsal bending by receiving excitatory inputs (from P cells) and output to excitatory motor neurons. They comment, "There are no functional connections between the interneurons."[14] This latter statement means, I presume, we must treat each interneuron as an independent part of the mechanisms. There, however, is a seeming paradox that needs to be explained away. Each interneuron has multiple inputs, and some of these inputs excite motor neurons that activate ventral bending (i.e. bending towards the stimulus, rather than the avoidance away from.) They try explaining this untoward (non teleological) activity in terms of its usefulness to other behaviors that the leech performs, e.g. swimming. What this implies is that some interneurons are part of a number of different systems having different goal states, which are all functioning together at any given time.

Similarly, every human sensory path carries multimodal information, part of which we neglect when we study the mechanisms of the visual system. This means that what counts as mechanism for a given end state is partial function of the purposes of the investigator, which is what I have called perspectival teleology. But only a partial function; the other constraints come from the world and more specifically our background knowledge about the world that constrain where the investigator looks, what she studies and tries to isolate and identify, what can be discovered, and, most importantly, what end or goal state is be chosen to be investigated.

This description is a pastiche of the discovery of mechanisms procedures that Lindley Darden and Carl Craver have begun to work out.[15] Notice though that the goal condition, aversive behavior, is taken as an unproblematic natural response, but it needs to be specified

[10] Cf. Churchland, P. and Sejnowski, T., *The Computational Brain*, Bradford Book, The MIT Press, Cambridge, MA, 1992, pp. 336f.

[11] For decomposition see Bechtel, W. and Richardson, R., *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Chapter 2.

[12] Churchland, P. and Sejnowski, T., *The Computational Brain*, p. 341.

[13] *The Computational Brain*, p. 342.

[14] Churchland, P. and Sejnowski, T., *The Computational Brain*, pp. 343-344.

[15] See Craver, C. F. and Darden, L., "Discovering Mechanisms in Neurobiology: The Case of Spatial Memory," in Machamer, P., Grush, R. and McLaughlin, P. (eds.), *Theory and Method in the Neurosciences*, University of Pittsburgh Press, Pittsburgh, 2001, pp. 112-137; and Craver, C. F., "Interlevel Experiments and Multilevel Mechanisms in the Neuroscience of Memory," *Philosophy of Science,* v. 69, (2002), pp. S83-S97.

in more detail. Aversive behavior was selected by the researcher because of its seemingly typical importance to organisms of many kinds in order to avoid pain or harm. Yet, there is no incompatibility in this case between being natural and being socially or personally chosen. It is both perspective teleology and natural teleology. The researcher selects the perspectival goal of the mechanism, although it is a naturally occurring outcome of a production in nature. It has a natural *telos*. The researcher also has to make many decisions as to what to include in the mechanism that produces the goal state, and these decisions too are constrained by what's there, what's happening, and what techniques she has available for investigation. One way to think about the need for such choices in naturalistic contexts is that there are no closed systems in nature, and so the researcher must put limits on them (close them) for explanatory and investigatory purposes. This echoes the rationale given for controlled experiments. Another dimension in the Churchland and Sejnowski leech case, the neurobiological information we have also constrains the building of a computational model. "The most basic lesson from LeechNet I and LeechNet II is that neural circuits could be constructed for performing the desired behavior, with components whose properties are consistent with the limited [anatomical and neurobiological] data available."[16]

There seem to be some limits. We noted the lack of functional connection among interneurons above. This provides us with anatomical constraint on what counts as the mechanism. More philosophically, we generally do not countenance mechanism explanations for "non natural," Cambridge complex properties (e.g., the object or event formed my mixing a martini today at 6 p.m. and Giovanni's turning off his Puccini recording yesterday.) We do not think we could ever find any productive connection among these activities. We have no useable criteria or background knowledge that places these events into a unified system. Our presumed knowledge at a time is used to specify what counts as a system as well as what system will be studied. If we wish to study some visual mechanism, our knowledge (borne, as Tom Kuhn said, from our training) of what constitutes the visual system puts boundaries on where and what kinds of mechanisms, entities and activities we may look for. Yet we may decide to question that knowledge.

What count as proper phenomena for explanations by mechanisms depend, most usually, on the criteria of importance that are in place in the discipline at a time. This means there is always an evaluative component operative in selection of goal states to be explained. Yet, there are physical systems with equilibrium states, where being in equilibrium or in homeostasis is the natural goal, and we seek to discover the mechanism by which such equilibrium is achieved. Here we might be tempted to say that this is a natural teleological system, the goal given by nature, and we just become interested in it. But if somehow we established there were no true equilibrium systems in nature, or that what we call an aversive response is really myriad of different systems operating in different ways, then it is we who would have made the mistake in identifying the phenomenon of our research and treating it as unitary.

There is another kind of mistake possible, which takes us back the reasons Churchland and Sejnowski gave for picking the leech as model organism. Their second criterion was ease of experimentation. This surely is a reasonable pragmatic constraint. But it can lead to error when the experimental paradigms we use are chosen primarily because we know they may be applied cleanly, rather than because they allow us to explore the phenomenon

---

[16] Churchland, P. and Sejnowski, T., *The Computational Brain*, p. 352.

of interest.[17] I am raising here questions about the validity of some quite reliable data. There is no time here to go into the problem in any depth, but consider Churchland and Sejnowski's characterization of the leech as "recognizing potential danger" when prodded. The behavior we see is that the leech bends away from the stimulus, and we call it aversive, and then gloss "aversive" as "response to potential danger." This description makes the behavior intelligible to us. Such a gloss is probably harmless enough in this case. But consider the experimental paradigms of learning used in neurobiology. They come in three types: repeated stimulation, classical conditioning and reward conditioning. There are problems internal to each of these paradigms that could lead one to question their use and, especially, the interpretation they provide for the data that is supposed to be learning. But let me raise a bigger question. At the time of the cognitive revolution (during the 60s, though one can find harbingers of this before, e.g. Bartlett 1932),[18] a major reason for shifting the paradigm from behaviorism was the inadequacy of just these three paradigms for learning. Could it be that neurobiologists use these paradigms because they can be applied despite the fact that much of the learning we are interested in, say in higher primates, cannot explained using only these limited kinds?

3. Reduction

All the above topics are related to old philosophical problem of reduction. In perhaps its classic form, Schaffner argued that uni-level reduction is the hallmark of a fully "clarified science,"[19] and that its realization requires the satisfaction of two specific conditions. First, causal generalizations contained in the *explanans* of a given theory must be completely reduced to terminology referring to in the processes occurring at one specific level of aggregation. Secondly, the *explanandum* must be uni-level in so far as it contains a generalization statement situated at the same or different level of aggregation as the *explanans*.

Scientists' use of the term "reduction" differs from the philosophers use. Scientists most often just mean that at least part of the explanation of a phenomenon is provided exhibiting the lower level mechanisms that show how that phenomenon is produced. The mechanisms are operating at different levels and, in some cases, are constitutive of the entities and activities at higher levels.

In fact, it seems that all or almost all mechanisms are multi-level. If this is so, the way philosophers see reduction, as reducing "higher" levels to a single lower level, is impossible. Some philosophers —Darden, Wimsatt, and Schaffner— have recognized this.[20]

[17] Aspects of this problem have been discussed in Weber, M., "Under the Lamp Post: Comments on Schaffner," in Machamer, P. K., Grush, R. and McLaughlin, P. (eds.), *Theory and Method in the Neurosciences*, University of Pittsburgh Press, Pittsburgh, 2001, pp. 231-249.

[18] Cf. Bartlett, F. C., *Remembering: A Study in Experimental and Social Psychology*, Cambridge University Press, Cambridge, 1932.

[19] Cf. Schaffner, K., *Discovery and Explanations in Biology and Medicine*, The University of Chicago Press, Chicago, 1993.

[20] Cf. Darden, L., "Discovering Mechanisms in Molecular Biology: Finding and Fixing Incompleteness and Incorrectness," in Schickore, J. and Steinle, F. (eds.), *Revisiting Discovery and Justification*, Max Planck Institute for History of Science, Berlin, 2002, pp. 143-154; Wimsatt, W. C., "Reductionism and its Heuristics: Making Methodological Reductionism Honest," paper for a conference on Reductionism, Institute Jean Nicod, Paris, November 2003. Proceedings forthcoming in *Synthese*; and Schaffner, K., *Discovery and Explanations in Biology and Medicine*, pp. 296-322.

4. Brain, Mind and Reduction

Since neuroscience is the science of the brain (loosely speaking) most people, it seems, are interested in what it has to tell about the mind/body, or perhaps better the mind/brain problem. The variations of the answers to this question are well known in outline to all who have studied philosophy. The major contrast class answers are, of course, materialism (or, nowadays, physicalism) and dualism. Materialism basically holds that mental phenomena do not exist, while dualism claims ontological independent existence of the two domains.

In contemporary times, the mind/brain conflict is usually stated in reductivist terms. Physicalists, in the extreme, claim that all mental locutions or entities can be eliminated or educed to the purely physical. The Paul and Patricia Churchland are qualified holders of this view, while John Bickle is its most strong and ardent supporter.[21] It is hard to find true philosophical dualists these days, though when the issue of the nature of consciousness or sometimes *qualia* of mental states is raised, there appear some who argue that consciousness or *qualia* cannot be reduced to physical states. [22]

As I shall speak more about below, much of the plausibility of the claims depends upon what is meant by "reduction" and/or "independence." Further, Since Spinoza there have been a variety of positions seeking to establish a *tertium modum* between these two extreme positions. The most favored forms today include supervenience, whereby a mental like state is supposed to be a quasi-causal consequence of a physical state, and so it is not truly independent, yet it is held to carry some sort of autonomy.[23] A more mystifying version is found in John Searle who basically hold that everything is physical and causal, including the mind, but that we need to talk about it independently since we cannot explain anything mental by anything physical.

I shall not here go into anything more about these BIG questions (for one reason, see Bechtel).[24] I shall have something to say, towards the end of this talk, about consciousness and I shall discuss reduction. I shall not say anything further about the mind/brain-body problem.

5. Information

A ubiquitous concept, *information*, is found throughout he neuroscience literature. It's use in science dates back to Claude E. Shannon and Warren Weaver,[25] who developed a mathematical concept of information, where information is simply a calculated probability that some bit of a structure at one end of channel (the sender) will end up at the other end, a receiver. This type of information cannot be used to talk about content, and has quite rigorous criteria for being applied to any communication system. Almost all of the time, when the term "information" is used in neuroscience it is not used in this mathematical sense.

---

[21] Cf. Bickle, J., *Psychoneural Reduction*, The MIT Press, Cambridge, MA, 1998; and Bickle, J. *Philosophy and Neuroscience: A Ruthlessly Reductive Account*, Kluwer, Dordrecht-Boston, 2003.

[22] Cf. Nagel, T., "What is it like to be a Bat," *The Philosophical Review,* v. 83, n. 4, (1974), pp. 435-450.

[23] Cf. Kim, J., *Mind in a Physical World*, Cambridge University Press, Cambridge, 1998.

[24] Bechtel, W., "Cognitive Neuroscience," in Machamer, P., Grush, R. and McLaughlin, P. (eds.), *Theory and Method in the Neurosciences*, University of Pittsburgh Press, Pittsburgh, 2001, pp. 81-111.

[25] Cf. Shannon, C. and Weaver, W., *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949, republished in paperback, 1963.

An attempt was made by Fred Dretske to develop a contentful concept of information.[26] Information was supposed to carry meaningful content. But it is universally agreed that it failed. Most recently, Ruth Millikan has just published a new philosophical account of information,[27] but it has not been evaluated as yet. "Information" has a bad philosophical reputation now. As my former colleague, Paul Griffiths said, information is a metaphor in search of theory. I think maybe it does not need a theory.

In neuroscience, mechanisms are said to carry information about various things. Neurotransmitters are said to carry information across synaptic clefts. More generally, perception systems are said to carry and process information about the environment. The "cognitivist" J. J. Gibson was one of the pioneers in this attempt to use information in a non-mathematical, useful way.[28]

Consider an ancient example: How does an eye become a seeing eye? Answer: by taking on the form of the object seen. The eye becomes informed by the object. This is a perfectly good concept of information, and though the theory of the seeing eye and the knowing mind has changed in many ways, the explanation for vision still proceeds in somewhat the same way. The eye [somehow to be specified in some degree of detail] picks up information about the world from an eye-world relation. So perhaps we ought to say about "information" what Tom Kuhn said about "paradigm," that it was a perfectly good word until he got hold of it. So too "information" was perfectly good, appropriately descriptive word before Shannon and Weaver mathematically defined it. There is nothing wrong with their probabilistic mathematical concept of information transmission. It just does not carry the semantic content that most people assume when using the concept.

Let us look at a house builder for an example of information and see how its relation to mechanisms and productive activities might help establish our intuitions. Let us say the builder starts with a blueprint, on which is recorded the information about what the owners have approved for their house. Therefore, in some clear sense the blueprint contains *information* insofar as it *represents* the desired house, i.e., the house actually built as desired is the *goal.* The blueprint serves the builder as a plan that if used properly can produce the goal. The builder uses the information in the blue print to build. The blueprint also serves as the part of the normative criteria for adequacy as to whether the plan has been executed properly.

In this use of information there is there is no signal nor other pattern passed on that remains intact from blue print to actual house during the period of construction. There is no channel along which anything flows. The information in the blue print is used to build the house, but it is not that the blueprint (as information) is somehow transduced and retains some structural aspect throughout the productive activities that bring the house into being. (Though, in some cases there may be such transductions.) Maybe a similar story that could be told about the biological case of DNA producing RNA, which then collects together specific strings of amino acids which make specific proteins on specific occasions, but here again the story gets complicated by the fact that any protein seemingly

[26] Cf. Dretske, F., *Knowledge and the Flow of Information*, The MIT Press, Cambridge, MA, 1981.

[27] Cf. Millikan, R., *Varieties of Meaning*, The MIT Press, Cambridge, MA, 2004.

[28] Cf. Gibson, J. J., *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, 1979 (reprinted in Lawrence Erlbaum Associates, 1987).

can be (is) made in number of different ways. Similar remarks may be told about synaptic transmission or about the visual system. There is nothing, no structure, that seems to remain intact and flow along a channel. It's always much more complicated.

We may identify a structure (or content) in the blueprint or, perhaps in a string of codons, in that this originating state, somehow to be explained in each case, represents the goal state or product, and which may be actively used by an agent, as instructions, in order to produce the goal state, though much more information and many more activities are needed to achieve the goal. The goal state is specified only partially by the information that is presented in the blueprint, i.e., in a quite limited way the blueprint presents aspects of what the house will look like when built. It is a limited 2-dimensional representation of what is to become a 3-dimensional object, and limited in that many details are missing, many skills needed to execute the blueprint are not described, etc. Conversely, because of the specificity of structures one can compare the house to the blueprint and see, in some respects, how well the builders followed their plan. But only in some respects; in just those respects where the information available in the blue print maybe compared with the information available from looking at the finished house. Similarly, information in DNA is information for the mRNA and it carries information about which amino acids, bases, nucleotides and other chemical structures may collaborate in the construction of a protein. But this is just to say we have a mechanism. The idea of information is not really doing any work.

So here is one moral: Information is a noun that describes "things" that provide information *for some use* by some one (or something). The major uses of information are as evidence for an inference, often inductive; a basis for a prediction; a basis for an expectation; or a basis that directs the subsequent behavior of a mechanism. The information in the blueprint to direct persons' (builders') actions, i.e., those actions are the tasks of construction necessary to build a house that looks like the one in the blueprint. The blue print portrays how to connect those things, but may not say what type of nails to use. Information is always information *for* an activity that will use it, and "use" means the information directs and controls the activity(ies) or production to some extent. Maybe we could say usefully, that information constrains the activities involved in the production.

Here is another moral: Things use information *for some purpose*. So information systems are always teleological in this sense. So information is always information *about*-where information "points towards some end."

We can distinguish two basic types of purposes: Information for *rational use* (as in inferences or as basis for interpretation), and information for *mechanical use* (for control or for production.)

The information in the blueprint is then *information about* the house. And this is true despite the fact that there is no message in the blue print (so whatever information is, it is not like letter or telegraph.) The encoding need not send signals, but the information is for the active producer and is used to produce by the producer. And we can check the end product, against the information in the starting conditions to see if it was successfully produced. This means information is always information about an end stage, and so is a relation between originating conditions and this termination stage, and the intervening activities are directed by the originating state towards the final product. So information is teleological.

Natural mechanisms may follow such selective and productive patterns also. Information, in this sense, is what directs a mechanism's activities of production such that they result in a selected end, where the end bears an appropriate specified relation to the beginning. So, as said above, it is helpful to think of information as constraining the possible outcomes by controlling the means and materials of production. In perception complex features of the environment are actively picked up by a perceiver and used by that perceiver to guide subsequent behavior. The information in a DNA segment is about what bases are to belong to the protein to be synthesized, and how those bases are to be arranged. But what makes it so is not the transmission of a signal. Just what information the DNA segment contains depends on the arrangements of the bases in its codons. DNA and mNRA segments feature different bases. As protein synthesis proceeds the patterns in which they are arranged are replaced by new patterns. Even if we could make sense out of the claim that the DNA contains information about the protein to be synthesized, no signal is transmitted from the DNA to the structures involved in the later stages of the synthesis.[29]

From this, we might say that the structure of the start up conditions selects the subsequent activities and entities that are responsible for bringing into existence termination conditions (which is our phenomenon of interest). One might be mislead by this way of speaking into thinking that intentional animism has crept in again, with the concept of responsibility and control and the need to attribute these to the proper entities and activities that we take to comprise the mechanism. However, this is a question of what is controlling the production of the end state and what relation that end state bears to the originating state. Often in such conditions, we are tempted to go outside of the mechanism and seek larger causes for and a greater specification of why the termination conditions are a preferred state. But this is true for any selective processes.[30] Noting this feature helps explain the desire to bring in evolution as a selection principle everywhere we find a mechanism because we think then we have a "big" reason for the goal. This is the job that God used to do. Unfortunately, in many cases evolution works no better than God did.

6. Computational Models

Computational models are often used to bridge the "gap" between neurobiology and gross physiology and behavior. Models present highly abstract sets of relations that assume that neurons in the brain may be described as functioning items in a digital mechanisms or networks, which if compounded and manipulated in the proper ways may show how possibly a brain can give rise to behavior. We refereed to one such model above, Churchland and Sejnowski's Leechnet I & II.[31] which was a computer model for the leech's dorsal bending.

Such models obtain their explanatory power because they may be used to mediate between known or assumed neurobiological data that serve as constraints for the model. Further constraints may come from location data and cognitive tasks. Such models may be used to direct research into the mechanisms implicated in some cognitive task.

[29] For details, see Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P., *The Molecular Biology of the Cell*, 4th ed., Garland, New York, 2002, Chapter 7.

[30] Cf. Machamer, P. K., "Teleology and Selective Processes," in Colodny, R. (ed.), *Logic, Laws, and Life: Some Philosophical Complications*, University of Pittsburgh Press, Pittsburgh, 1977, pp. 129-142.

[31] On Churchland and Sejnowski's Leechnet I & II, cf. Churchland, P. S. and Sejnowski, T., *The Computational Brain*, passim.

7. Knowledge, representation and memory

An organism is said to have knowledge when it acquires and maintains information from its environment or some other source (including itself). I would add the additional condition, that it must be able to use this information in a normatively appropriate way. In any discourse on knowledge there is a set of topics that both philosophers and scientists must address. Basically these boil down to (1) acquisition or learning, (2) maintenance or memory systems, (3) use and (4) normative constraints that obtain for what is used to count as knowledge.[32]

Many studies in neuroscience focus on how learning occurs and how memory is maintained. It makes little sense to speak of acquisition or learning if what is acquired if what is learned is not represented somehow in some of the organism's systems. So, any knowledge, the product of learning mechanisms, must be "represented" in one or more of the memory systems, whether sitting calmly as a trace, as a somewhat permanent systemic modification of some kind (e.g., perhaps as a set of long term potentiations at a micro-level or as a revalued cognitive or neural network at a higher systemic or computational level) or in motor systems representing a tendency towards or an ability to perform certain actions. This memorial maintenance requirement does not entail an identity between knowledge and the representation (or the content of the representation), for there is much evidence that, at least some kinds of, knowledge as expressed under certain task conditions comes from constructing or reconstructing by adding to some stored (represented) elements. The most usual situations of this kind are recall tasks. Think about the windows on the ground floor of your house. Count them. The phenomenology of this task suggests strongly that you are constructing this visual image as you are proceeding through the house, counting as you go. Yet by performing this counting task, you become aware that you know the number of windows.

It is fairly traditional in today's neuroscience to identify kinds of knowledge with kinds of representations in different kinds of memory systems. The classic cognitive taxonomy for the different systems were: iconic storage; short term memory; and long term memory.[33]

In neuroscience the first representations (formerly iconic) are present in the various subsystems of sensory systems. In perception, information is extracted from the environment, and processed through various hierarchies of the perceptual systems. These have outputs that, in cases where long term, declarative memories will occur, are passed into the working memory system (formerly short term memory). Working memory is assumed to be at least partially located in the hippocampus. Working memory is where the binding together of various inputs from the different sensory systems occurs. From there, perhaps when some sort of threshold has been attained, the information goes into one of three or four different types of long term systems: (1) Declarative (long term, encoded in "bits," or concepts, categories); (2) Semantic (linguistic; words, propositions); (3) Episodic (autobiographical, relation to "I"; personal experiences); and (4) Spatial (spatial relations specifying where (a point of view or location system). The shift from working memory into a long term system calls for *memory consolidation*.

[32] A version of part of this section looking at social aspects of learning, memory and knowledge was published in Machamer, P. K. and Osbeck, L., "The Social in the Epistemic," in Machamer, P. K. and Wolters, G. (eds.), *Science, Values and Objectivity*, University of Pittsburgh Press, Pittsburgh, 2004, pp. 78-89.

[33] Cf. Neisser, U., *Cognitive Psychology*, Appleton Century Crofts, N. York, 1967.

There is also, most importantly, another memory system labeled (5) Procedural Memory. There are probably a number of different systems grouped under this rubric. Procedural memories are formed from having learned how to perform different sorts of actions or skills. They too must be represented in the bodily system, but do not at least usually seem to go through the hippocampal system of working memory. There are also (many) (6) emotional systems. Finally, some sorts of (7) attention systems have to be considered. Exactly where all these systems are located in brain is under some debate. Also there are questions as to whether the types of systems I have just enumerated are the "right" ones.

Very often network connections in long term memory are modeled as networks of information, taken as categorical structures or schemata, into which inputs fit and which change in interconnective relational weight as learning occurs. The connections in the network then represent inference possibilities (lines of inference) and/ or expectations that may follow from a particular categorization. Also often the memory schema relates various appropriate actions or motor skills. There are also "off-line" networks, which are representation systems (perhaps like imagination) that allow us to process information and draw conclusions without actually acting.

A few philosophical points need to be drawn from the above descriptions of memory or knowledge systems. First, from the detail of these different systems it should be clear that the philosopher's distinction between *knowing how* and *knowing that* is quite inadequate. Philosophers need to think of many more kinds of knowledge than they have heretofore considered.

Second, while not directly the concern of most neuroscientists, it is of philosophical note that in order for any procedural or declarative or otherwise "encoded" in memory to count as knowledge, it must be presupposed that there are shared social norms of appropriate or correct *application* of what has been acquired. These application norms apply to the uses of what has been internalized and function as criteria as to whether one has acquired knowledge at all. To count as knowledge, whatever is learned and stored must be able to be used by the knower. It is epistemologically insufficient to theorize knowledge only in terms of acquisition and representation. One must also consider how people *exhibit* that learning has taken place through appropriate or correct action (including correct uses of language). For this reason one major research interest in neuroscience is the inextricable connection between what used to be called afferent and efferent systems, or perception, memory and motor systems. But this very terminology fails to take into account that in some sense in perception, memory and knowledge, there is a larger system whose mechanisms need to become the focus of more research.

Knowledge, in part, is constituted by action, broadly defined, e.g. action here includes appropriate judgment or use of words. And it is inseparably so constituted. That is, knowledge always has a constitutive practical part. We might say that every bit of theoretical knowledge is tied inextricably with practical reason, a thesis that dates back to Aristotle.[34] Importantly, these knowledge acts occur in a social space wherein their application is publicly judged as appropriate or adequate (or even correct). One way to see

[34] Cf. Aristotle, *Nichomachean Ethics*, edited with translation by H. Rackham, Harvard University Press, Cambridge, 1934, book VI, 7, 1141b13ff.

this is to note that truth conditions, at least in practice, are use conditions, and even those "norms set by the world" (empirical constraints) structure human practices concerning correct or acceptable use, e.g., of descriptions of use of instruments or even the right ways to fix a car. Human practices are social practices, and the way we humans use truth as a criterion is determined not only by the world, but also by those traditions we have of inquiring about the world, of assessing the legitimacy of descriptions and claims made about the world, and of evaluating actions performed in the world.

As extrinsic epistemologists have stressed there is a need to refocus philosophical concern to concentrate on *reliability* (rather than truth.) From our discussion above we may infer that the reliability of memory entails answers about reliability of knowledge.

One way to begin to think about reliability and its connection to truth is, what are the conditions for "reliable" or appropriate assertions (or other kinds of speech acts)? That is, what conditions must obtain or what presuppositions must be fulfilled in order to make a speech act reliable or appropriate? What are the criteria of success for such speech acts? And how are such criteria established? Validated? And changed? More generally what evaluative criteria apply to actions? Speech acts are actions. How do actions show yourself, you who perform them, that they are effective? Appropriate? These are different. You may effectively insult someone, but it maybe highly inappropriate. On what grounds are they judged to be appropriate by other people?

8. Consciousness: What is the problem?

The problem of consciousness is that it is not one problem. Many things have been discussed under the ideas of consciousness. We have little time remaining so I shall be merely provocatively assertive here. The big question about consciousness can be put: why are certain forms of information used be people displayed in the qualitative form of conscious experience. Put slightly differently, what is there that a person can do if one perceptual systems display information in this modally specific qualitative ways, and how can this information be used in ways that it cannot if it were made available in some other form? One general answer to this question is called the *conscious workspace hypotheses*. Yet another aspect that gets raised in what evolutionary advantage much such a system have such that it would have been selected for? Finally, there are some neuroscientists who have been working on trying to establish the mechanisms that produce this form of information display. But these issues take us into new realms, and cannot explored further here.

9. Selected Bibliography (and for further reading)

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P., *The Molecular Biology of the Cell*, 4th ed., Garland, New York, 2002.

Anscombe, G. E. M., "Causality and Determination," Inaugural lecture at the University of Cambridge, 1971; reprinted in Anscombe, G. E. M., *Metaphysics and the Philosophy of Mind, The Collected Philosophical Papers,* 1981, vol. 2., University of Minnesota Press, Minneapolis, pp. 133-147.

Aristotle, *Nichomachean Ethics*, edited with translation by H. Rackham, Harvard University Press, Cambridge, 1934.

Bartlett, F. C., *Remembering: A Study in Experimental and Social Psychology*, Cambridge University Press, Cambridge, 1932.

Bechtel, W., *Philosophy of Science. An Overview for Cognitive Science*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1988.

Bechtel, W. and Richardson, R., *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Princeton University Press, Princeton, 1992.

Bechtel, W., "Cognitive Neuroscience," in Machamer, P., Grush, R. and McLaughlin, P. (eds.), *Theory and Method in the Neurosciences*, University of Pittsburgh Press, Pittsburgh, 2001, pp. 81-111.

Bickle, J., *Psychoneural Reduction*, The MIT Press, Cambridge, MA, 1998.

Bickle, J. *Philosophy and Neuroscience: A Ruthlessly Reductive Account*, Kluwer, Dordrecht-Boston, 2003.

Bogen, J. and Woodward, J., "Saving the Phenomena," *The Philosophical Review*, v. 97, (1988), pp. 303-352.

Chalmers, D., *The Conscious Mind*, Oxford University Press, New York, 1996.

Churchland, P. S., *Neurophilosophy: Towards a Unified Science of the Mind-Brain*, The MIT Press, Cambridge, MA, 1986.

Churchland, P. M. and Churchland, P. S., "Intertheoretic Reduction: A Neuroscientist's Field Guide," *Seminars in the Neurosciences*, v. 2, (1991), pp. 249-256.

Churchland, P. S. and Sejnowski, T., *The Computational Brain*, Bradford Book, The MIT Press, Cambridge, MA, 1992.

Craver, C. F., *Neural Mechanisms: On the Structure, Function, and Development of Theories in Neurobiology*, Ph.D. Dissertation, University of Pittsburgh, Pittsburgh, PA, 1998.

Craver, C. F., "Role, Mechanisms, and Hierarchy," *Philosophy of Science*, v. 68, (2001), pp. 53-74.

Craver, C. F. and Darden, L., "Discovering Mechanisms in Neurobiology: The Case of Spatial Memory," in Machamer, P., Grush, R. and McLaughlin, P. (eds.), *Theory and Method in the Neurosciences*, University of Pittsburgh Press, Pittsburgh, 2001, pp. 112-137.

Craver, C. F., "Interlevel Experiments and Multilevel Mechanisms in the Neuroscience of Memory," *Philosophy of Science,* v. 69, (2002), pp. S83-S97.

Craver, C. F., *Explaining the Brain: What a Science of the Mind-Brain Could Be*, Oxford University Press, Oxford, forthcoming.

Darden, L., "Discovering Mechanisms in Molecular Biology: Finding and Fixing Incompleteness and Incorrectness," in Schickore, J. and Steinle, F. (eds.), *Revisiting Discovery and Justification*, Max Planck Institute for History of Science, Berlin, 2002, pp. 143-154.

Dennett, D., *The Intentional Stance*, The MIT Press, Cambridge, MA, 1987.

Dretske, F., *Knowledge and the Flow of Information*, The MIT Press, Cambridge, MA, 1981.

Fodor, J., "Special Sciences (Or: The Disunity Of Science As A Working Hypothesis)," *Synthese*, v. 28, (1974), pp. 97-115.

Fodor, J., "Special Sciences: Still Autonomous After All These Years," in Toberlin, J. (ed.), *Philosophical Perspectives 11: Mind, Causation, and World*, Blackwell, Boston, 1997, pp. 149-163.

Glennan, S., "Mechanisms and the Nature of Causation," *Erkenntnis*, v. 44, (1996), pp. 49-71.

Glennan, S., "Rethinking Mechanical Explanation," *Philosophy of Science*, v. 69, (2002), pp. S342-S353.

Gibson, J. J., *The Senses Considered as Perceptual Systems*, Houghton Mifflin, Boston, 1966.

Gibson, J. J., *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, 1979 (reprinted in Lawrence Erlbaum Associates, 1987).

Griffiths, P. and Stotz, K., "Representing Genes," *www.pitt.edu/~kstotz/genes/genes.html*.

Hempel, C. G., "The Logic of Functional Analysis," in Gross, Ll. (ed.), *Symposium on Sociological Theory*, P. Row, Evanston, IL, 1959, pp. 271-307; reprinted in Hempel, C. G., *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*, Free Press, N. York, 1965, pp. 297-330.

Kim, J., *Mind in a Physical World*, Cambridge University Press, Cambridge, 1998.

Machamer, P. K., "Gibson and the Conditions of Perception," in Machamer, P. K. and Turnbull, R. G. (eds.), *Perception: Historical and Philosophical Studies*, Ohio State University Press, Columbus, OH, 1975, pp. 435-466.

Machamer, P. K., "Teleology and Selective Processes," in Colodny, R. (ed.), *Logic, Laws, and Life: Some Philosophical Complications*, University of Pittsburgh Press, Pittsburgh, 1977, pp. 129-142.

Machamer, P. K., Darden, L. and Craver, C. F., "Thinking About Mechanisms," *Philosophy of Science* v. 67, n. 1, (2000), pp. 1-25.

Machamer, P. K., Grush, R. and McLaughlin, P. (eds.), *Theory and Method in the Neurosciences*, University of Pittsburgh Press, Pittsburgh, 2001.

Machamer, P. K. and Osbeck, L., "The Social in the Epistemic," in Machamer, P. K. and Wolters, G. (eds.), *Science, Values and Objectivity*, University of Pittsburgh Press, Pittsburgh, 2004, pp. 78-89.

Millikan, R., *Varieties of Meaning*, The MIT Press, Cambridge, MA, 2004.

Mitchell, S. D., *Biological Complexity and Integrative Pluralism* (*Cambridge Studies in Philosophy and Biology*), Cambridge University Press, Cambridge, 2003.

Nagel, T., "What is it like to be a Bat," *The Philosophical Review,* v. 83, n. 4, (1974), pp. 435-450.

Neisser, U., *Cognitive Psychology*, Appleton Century Crofts, N. York, 1967.

Richardson, R., "Discussion: How Not to Reduce a Functional Psychology," *Philosophy of Science*, v. 49, (1982), pp. 125-137.

Richardson, R., "Cognitive Science and Neuroscience: New-Wave Reductionism," *Philosophical Psychology*, v. 12, n. 3, (1999), pp. 297-307.

Robinson, H., "Dualism," in Stich, S. and Warfield, T. (eds.), *The Blackwell Guide to Philosophy of Mind*, Blackwell, Oxford, 2003, pp. 85-101.

Robinson, H., "Dualism," *The Stanford Encyclopedia of Philosophy (Fall 2003 Edition)*, Zalta, E. N. (ed.), URL =<http://plato.stanford.edu/archives/fall2003/entries/dualism/>.

SARKAR, S., "Models of Reduction and Categories of Reductionism," *Synthese*, v. 91, n. 3, (1992), pp. 167-194.

SCHAFFNER, K., "Approaches to Reductionism," *Philosophy of Science*, v. 34, (1967), pp. 137-147.

SCHAFFNER, K., "Theory Structure, Reduction and Disciplinary Integration in Biology," *Biology and Philosophy*, v. 8, n. 3, (1993), pp. 319-347.

SCHAFFNER, K., *Discovery and Explanations in Biology and Medicine*, The University of Chicago Press, Chicago, 1993.

SCHAFFNER, K., "Interactions Among Theory, Experiment, and Technology in Molecular Biology," *Proceedings of the Biennial Meetings of the Philosophy of Science Association*, v. 2, (1994), pp. 192-205.

SCHAFFNER, K., *Reductionism and Determinism in Human Genetics: Lessons from Simple Organisms*, University of Notre Dame Press, Notre Dame, IN, 1995.

SCHOUTEN, M. and LOREEN-DE-JONG, H., "Reduction, Elimination, and Levels: The Case of the LTP-Learning Link," *Philosophical Psychology*, v. 12, n. 3, (1999), pp. 237-262.

SHANNON, C. and WEAVER, W., *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949, republished in paperback, 1963.

SMART, J. J. C., "Sensations and Brain Processes," *Philosophical Review*, v. 68, (1959), pp. 141-156.

WEBER, M., "Under the Lamp Post: Comments on Schaffner," in MACHAMER, P. K., GRUSH, R., and MCLAUGHLIN, P. (eds.), *Theory and Method in the Neurosciences*, University of Pittsburgh Press, Pittsburgh, 2001, pp. 231-249.

WIMSATT, W., "The Ontology of Complex Systems: Levels of Organization, Perspectives and Causal Thickets," *Canadian Journal of Philosophy of Science*, Supplementary Volume 20, (1994), pp. 207-274.

WIMSATT, W. C., "Reductionism and its Heuristics: Making Methodological Reductionism Honest," paper for a conference on *Reductionism*, Institute Jean Nicod, Paris, November 2003. Proceedings forthcoming in *Synthese*.

WOODWARD, J., "Data and Phenomena," *Synthese*, v. 79, (1989), pp. 393-472.

Education, the Brain and Behavior: Reflections on Today's Psychology

Jose Sanmartin

This chapter analyzes an interactionist alternative, based on social-cognitive theory, to the environmentalist or biologist explanations for some aspects of human behavior, in particular, jealousy-related violence.

1. On Environmentalism

Environmentalism has systematically attempted to explain human behavior and some of its most striking expressions, such as violence, as the product of certain environmental inputs and, especially, as the result of certain forms of group organization. This current was particularly popular under Marxism and behaviorism, which considered a person and their biology to be mere clay, to be molded by social inputs.

Thus, strictly speaking, Marxism does not acknowledge the existence of mental or personality disorders; it knows only social diseases. There are no possible organic causes for the complete lack of empathy and remorse with which psychopaths commit the most horrendous crimes, even deriving pleasure from them. A psychopath's biology is irrelevant. The fault lies elsewhere, in the sick society that produces such monsters, a society whose worst ills, such as violence, are caused by the existence of private property. After all, weren't humans peaceful during their hunting-gathering days, when cooperation and sharing were the key elements of group life? With the appearance of agriculture, however, paradise was supposedly lost. The way to restore peace, then, is to recreate the pre-Neolithic conditions of social life, making completely impossible aberrations such as the capturing, torturing, raping, killing and even eating of other human beings.

The establishment of socialist society would create those ideal conditions, and psychopaths, who would be better designated as "sociopaths," would no longer exist. This is what made it possible for Soviet authorities in the Rostov region to initially deny the obvious when, in the late 70s, they started finding an increasing number of brutally mutilated dead bodies of women, boys and girls. This was simply impossible; it was something typical of capitalism, not true socialism.

Denial grew even stronger when an unassuming civil servant, who was also a member of the Communist Party, was arrested in connection with these crimes. As if this was not enough, nature itself conspired to make things even more complicated, for this individual, Andrei R. Chikatilo, possessed a rare characteristic: his blood type did not match that of his secretions, causing the forensic tests to discard him as a suspect, since his samples did not match those found on the victims. Socialist conscience could breathe easily again. Not until perestroika was Chikatilo rearrested and accused of 36 murders, to which he voluntarily added 18 more. True socialism did produce monsters, after all...

This story serves as a small illustration of the nefarious consequences of extremisms. They deform the lenses through which we view the world and can take us to positions as ridiculous as the one just described, which is in no way unique to environmentalism, as I will now show.

2. Wilson and Sociobiology

I still remember the afternoon that I bought Edward Wilson's *Sociobiology* in a Munich bookstore. I think it was in March 1976, during my first stay in Germany. I have to admit that books such as this one opened my eyes to much more interesting matters than the model theories that I was somewhat obsessively studying in those days.

Wilson's book is both complex and ambiguous. It is no surprise to me, then, that many people do not use Wilson's book when discussing sociobiology. Rather, they use their own popularizations, in which some very important nuances related to human beings have disappeared completely. Specifically, Wilson states that his analyses of the human being are highly conjectural, thereby creating a wide gap between his observations of non-human animals and his speculations about human beings. Only in the case of non-human animals is behavior primarily the product of genes.

Wilson's voice of caution regarding human behavior was soon drowned out by the storm of media opinion his theories unleashed and by a somewhat frivolous philosophy. An idea that became particularly popular was that genes are the primary reason for such undesirable behavior as violence. And since violence can supposedly lead us toward extinction, any technology that influences its underlying genes and saves us from ourselves is considered highly welcome.

There is a special type of violence that has attracted the attention of many followers of this simple but dominant view of sociobiology: the abuse that women are subjected to by men. For these sociobiologists, violence is a man's tool for dominating a woman he is intimate with, the aim being to ensure sexual exclusiveness within the framework of incompatible reproductive strategies. On the one hand, men want to copulate with as many women as possible in order to maximize their contribution to the gene pool, at the same time excluding all possible rivals. In other words, they want harems. On the other hand, women want to maximize their chances of reproductive success by choosing a mate who has access to sufficient resources and is willing to share them with her *exclusively*, because they have to gestate and rear their offspring and these activities require a sizeable investment of resources. These two strategies are clearly incompatible: men cannot satisfy their short-term desires without harming women's long-term interests.

Some evolutionary psychologists have subsequently gone down the same path, David Buss foremost among them.

3. Jealousy in Evolutionary Psychology

Evolutionary psychology is the discipline that identifies those psychological mechanisms behind human behavior that are the product of evolution. I personally believe this to be a worthwhile and necessary enterprise. The problem, however, is that evolutionary psychologists maintain that all human behavior is **determined** by these mechanisms.

This evolutionarily determined behavior includes our sexual strategies, which are supposed to simply be adaptive solutions to mating problems. They do not require any conscious planning. It is even said that with sexual strategies it is like with a concert pianist's practice: the less one's consciousness intervenes, the better.

For evolutionary psychology, these sexual strategies include the choosing, attracting, keeping and replacing of a mate. I repeat. For evolutionary psychology, these activities are for the most part, if not completely, based on hereditary mechanisms, i.e., they are genetic in origin. So, even when a woman thinks she consciously chooses a mate for economic reasons, she is really following her genetic programming. To the evolutionary psychologist, she chooses a mate who signals his willingness to commit to her. Why? Because she has to internally gestate and then raise her offspring. These tasks require resources, and she needs a partner by her side who can provide these resources, who is willing to invest in her. In this sense, the biologically most effective women since the Pleistocene have been those whose biology was not fooled by men who were only interested in copulating without any strings attached.

As evolutionary psychology has it, this is precisely men's sexual strategy: copulating without committing and, when making a commitment, ensuring maximum fidelity from the partner. Men want women to be faithful, because unlike them, they cannot be sure of their fatherhood. A man, therefore, does not want to invest resources in an unfaithful woman, because he could be raising someone else's progeny at the expense of his own biological fitness.

In summary, according to evolutionary psychology, women look for commitment in their partners, while men evade commitment. So, we have a conflict of sexual strategies on our hands. This conflict continues when a man commits himself, because even then there is always an element of uncertainty surrounding his fatherhood. However, evolution has supposedly again come to the rescue, providing men and women with emotions that are unconsciously triggered when certain alarm signs appear. These are the emotions that constitute "sexual jealousy."

In the same way that men and women have different sexual strategies, they also experience sexual jealousy differently. Women become especially jealous when their partners show signs of emotional infidelity, i.e., when they are in love with somebody else. And this is exactly how things should be, according to evolutionary psychologists, because if there is anything that seriously harms a woman's biological fitness, it is having a partner who falls in love with another woman and stops investing resources in her. In short, women are biologically prepared to be less tolerant of emotional than sexual infidelity. The former usually entails abandonment, the latter does not. Men, in contrast, experience sexual jealousy especially in response to sexual infidelity, because our species' internal fertilization always leaves some doubt concerning fatherhood. Again, this was to be expected biologically, because, following these conceptions, the worst thing that can happen to a man is for his partner to gestate and raise another man's offspring.

Evolutionary psychology considers that men and women's different forms of sexual jealousy are the result of different innate modules. An innate module is defined as a set of philogenetically acquired brain circuits. In the case that interests us, these modules contain the instructions for jealousy. Of course, all of this starts from the proposition that, somewhere in the distant past, very probably during the Pleistocene, our ancestors' increased their biological fitness by feeling jealousy.

Sexual jealousy is triggered by certain stimuli that threaten an intimate relationship. As previously explained, in the case of men, the stimuli are related with sexual infidelity in the partner, and in the case of women, with emotional infidelity. This hypothesis is assumed to be correct because multiple studies conducted among college students, using the self-report format, have confirmed it. Now, these studies used forced-choice questions, i.e., the participating students had to choose one of two mutually exclusive alternatives: they were either jealous because of a supposed emotional infidelity, or because of a supposed sexual infidelity. The results of these studies show that the overwhelming majority of women (around 70%) are more hurt and worried about emotional infidelity, while many men (around 50%) have more of an issue with sexual infidelity.

Sometimes a woman's sexual infidelity can be such an extreme provocation that some evolutionary psychologists such as David Buss consider that even a *reasonable man* may legitimately respond with lethal violence. It is hard to reconcile male reproductive success with as inadaptive an act as the destruction of a key reproductive resource. However, David Buss claims that, from a reproductive point of view, killing one's own mate did not necessarily always have to be harmful throughout human evolution. First of all, when a woman abandons her mate, the mate in question does not only lose a reproductive resource, but that resource may very well end up in a competitor's hands, which would be doubly harmful to his reproductive success. In the second place, a man who consents to being cheated on will lose face and suffer ridicule if he does not compensate in some way. If he fails to do so, his future chances of having access to reproductive resources (another mate) may be seriously reduced. In short, authors like David Buss believe that the thought and act of killing were in all likelihood adaptive at some point in human history and are, therefore, a part of our evolutionary mechanisms. In other words, men are killing their partners today because in the past it improved their ancestors' biological fitness.

The conflict between men and women's sexual strategies does not only play itself out in the choice of a suitable mate, but also in the attempts to retain that mate. Here, once more, men and women display different behaviors, violence again being a typically male behavior, according to evolutionary psychology. In this case, violence is more likely to be psychological than physical, i.e., it will tend to undervalue or intimidate women, mainly through verbal means. Denigrating remarks concerning a woman's physical appearance are one way to tilt the balance of power in a man's favor. Sexually disparaging insults achieve submission through feelings of shame. Insults also help to ensure sexual exclusivity: if a man subdues a woman psychologically and socially, and restricts her possibilities of being unfaithful, he can pass on his genes with full confidence in his fatherhood.

4. On Social-Cognitive Theory

I disagree with evolutionary psychology's position on sexual jealousy and other characteristics of human behavior. Far be it from me to deny that there is a biological basis to sexual jealousy and to the majority of those other characteristics. However, I prefer to maintain a non-biologistic position in this context, siding, for example, with the "social-cognitive theory" of emotions.

According to the social-cognitive point of view, emotions have two basic components: a primary one, engraved in the brain by evolution, and a secondary one reflecting the contents of socialization. Depending on the contents acquired during socialization, the secondary component confers different meanings to the primary one.

Thus, as pointed out by Christine Harris, in the case of jealousy, the primary component probably had nothing to do with partner selection and romantic relations, but rather with sibling relations, with obtaining resources from parents. Jealousy has even been observed between babies that were just a few months old. Obviously, jealousy at such an early age cannot be influenced by reasoning or, at least, not by complex reasoning. It can, therefore, perfectly be called an instinctive reaction.

This innate trait, however, comes under the influence of the various contents acquired in the process of socialization. Depending on these contents, the instinctive trait will appear in different forms and in different contexts (between friends, between partners, etc.). For example, if someone receives a sexist socialization, if they learn to see their partner not as biologically different but as socially inferior, if they learn to depersonalize their partner to the point of viewing them as an object, as a piece of property, then it is obvious that the biologically normal expression of jealousy transforms into a socially abnormal form of jealousy. This type of pathological jealousy is more a desire to possess than a show of romantic passion. It is a possessive, not passionate, type of jealousy.

The contents of the socialization process are not engraved in philogenetically acquired neural circuits. They are acquired after birth and whether the neural circuits containing them express themselves or not depends on socialization itself. At least, this is what modern neuroscience allows me to state, and I wish to make it clear that I do not aim to undervalue the role of biology in any way; I just want to create a balance between the role of biology and society in the origin and development of the oddest animal currently walking the face of the Earth: Man.

5. Man as a Cultural Animal

Like other animals, Man changes and survives through biological evolution. However, as Ortega y Gasset pointed out, Man differs from other animals in that he has been more interested in living well than in just surviving. By "living well" I mean living without feeling the pressure of the necessities that have turned Man into a needy creature.

To achieve ever higher standards of wellbeing, Man has created a kind of artificial environment that is superimposed on the natural environment. This kind of supranature is composed of technical elements such as instruments, tools, machines, constructions and diverse forms of social organization that enable Man to meet his immediate needs, such as having food available whenever he feels like eating, eliminating cold in winter and heat in summer (at least, within certain spaces), bridging large distances, etc.

This artificial environment is, obviously, the product of **culture** and encompasses much more than just material and technical things. It contains and, at the same time, is the result of, certain ideas and beliefs. This network of ideas and beliefs is used, with or without the help of technical instruments, to **process** everything: from one's own self to relations with others and the environment.

6. The Human Brain

The brain is our big information-processing unit. But it is not a one-piece unit, but a collection of different components that, albeit interconnected, have very different origins and functions.

Although the term "limbic brain" has been widely questioned, it is undeniable that our brain contains a series of structures, directly linked to emotions, which act automatically in the presence of certain stimuli. For example, the facial features of a baby usually elicit tenderness in adults.

In short, the hypothesis about the existence of innate emotions has been proven beyond all doubt. Science has gradually identified the different brain circuits in which those emotions are inscribed.

Some complex behaviors are made up of several emotions. One of them is aggressive behavior, which is coordinated by a structure called the amygdala, and consists of a series of physiological reactions that are mainly produced by various nuclei located in the brain stem, hypothalamus and hippocampus.

When receiving certain sensory inputs, the amygdala triggers a series of expanding chain reactions. It is important to note that these reactions are instinctive. They take place without the person knowing why they occur. They are adaptive, however, first and foremost because they are crucial in situations of real or apparent danger. For example, I am walking through a forest. Suddenly, I freeze. I start sweating, my heartbeat and breath accelerate, my pupils widen and, in general, my senses sharpen. Why? Something on the ground, covered by leaves, attracted my attention. Now I see with relief that it was just a twisted stick, but it could have been a snake, and my instinctive reactions primed me to react unconsciously, without thinking. This particular set of instinctive reactions constitutes a special type of aggressive behavior called "defensive aggressiveness."

From a biological point of view, the transition from my initial instinctive reactions (freezing, sweating, etc.) to my posterior relief is a giant leap. If I can see clearly that the stick is just a stick, it is because the same sensory input that went directly to my amygdala also took a different and much longer route to my occipital cortex, where it was processed until the possible threat was positively identified as a stick, not a snake. This processed visual input was then sent back to the amygdala, which ceased sending out the instructions that kept my instinctive reactions going, thereby reestablishing tranquility.

7. The Prefrontal Cortex

The amygdala did not only receive a processed image from the occipital cortex, however. The different parts of the prefrontal cortex also sent analyses, comparisons and assessments of the situation, which were responsible for enhancing or reducing the until then automatically produced aggressive response. They were the result of the ideas and beliefs contained in the different parts of the prefrontal cortex, making this part of the brain the great regulator of aggressiveness.

Aggressiveness stops being mere aggressiveness and turns into violence when the prefrontal cortex enhances the aggressive response and confers intentionality, making violence the domain of Man (at least *for the most part*).

Furthermore, violence features the two key components of social-cognitive theory. First the primary component, aggressiveness, which is inscribed in philogenetically acquired brain circuits. The other, the secondary component, is made up of the ideas and beliefs that influence aggressiveness, orienting it, giving it meaning, in short, bringing it under the control of culture, of acquired aspects. And it could hardly be any different, because the neural circuits producing those ideas and beliefs in the different areas of the prefrontal cortex are not biologically inherited, or, at least, not completely. We now know that our experiences starting at childbirth (and maybe even before) can shape brain circuits or, alternatively, cause the brain to give priority to some circuits over others. This is why psychotherapy often fails to yield the desired results in the short term, because it is asked to perform the often very complicated biological feat of disconnecting brain circuits that have been favored or constructed by experience.

8. Final Remarks

When trying to explain or understand something, Man has a marked tendency to dichotomize. This is clearly shown by the history of philosophical thought, where one has to be either a rationalist or empiricist, an idealist or materialist, etc. Sometimes this tendency to dichotomize reaches pathological extremes, as is often the case in violent people. They routinely divide the world into "them" and "us," blaming others for all the ills that befall them.

This dichotomizing tendency has grown deep roots in the philosophy of biology and animal behavior research. Both fields have historically been battlegrounds for biologists and environmentalists, both sides offering attractively simple explanations for something as complicated as human behavior. The problem is that sometimes their explanations, more than simple, are simplistic.

Man is a very complex being and so is his behavior. Only if we dare to leave our engrained dichotomy behind, with the help of neuroscience, can we create the right conditions for discussing human behavior without prejudices, as what it is: a type of behavior resulting from a thick layer of culture covering a necessary but insufficient biological basis. This layer of culture hides and even reconfigures the biological basis, adding an aspect of intentionality that biology lacks of itself. This occurs with violence, and the same thing happens with jealousy. Biological jealousy is one thing, but sexual jealousy is something completely different: the primary biological component is reinterpreted and reoriented by the products of socialization. When one understands this, one also understands that men abuse their partners not because they are driven by biology, but because —to them— the stereotypes and beliefs acquired during their socialization *justify* their violent and sometimes even lethal behavior toward their partners. In summary, men do not consider their partners as property because of their genes, but because of the ideas and beliefs with which they reinterpret their biological jealousy.

9. Bibliography

Buss, D., *The Dangerous Passion: Why Jealousy is as Necessary as Love and Sex*, Free Press, New York, 2000.

Buss, D., "Human Mate Guarding," *Neuroendocrinology Letters*, v. 23, suppl. 4, (2002), pp. 23-29.

BUUNK, B., ANGLEITNER, A., OUBAID, V. and BUSS, D., "Sex Differences in Jealousy in Evolutionary and Cultural Perspective," *Psychological Science*; v. 7, n. 6, (1996), pp. 359-363.

DAMASIO, A., *Descartes's Error. Reason and the Human Brain*, A Grosset/Putnam Book, G. P. Putnam's Sons, New York, 1994. Spanish translation by Joandomènech Ros: *El error de Descartes*, Editorial Crítica, Barcelona, 1996.

DAWKINS, R., *The Blind Watchmaker*, W. E. Norton, London, 1986. Spanish translation by Manuel Arroyo Fernández: *El relojero ciego*, RBA Coleccionables, Barcelona, 1993.

DAWKINS, R. *The Selfish Gene*, Oxford University Press, Oxford, 1976. Spanish translation by Juana Robles Suárez and José Manuel Tola Alonso: *El gen egoísta*, Salvat Editores, Barcelona, 2000.

EIBL-EIBESFEDLT, I., *Grundriss der Vergleichenden Verhaltensforschung*, Piper and Co., München, 1969. Spanish translation by Margarida Costa: *Etología: introducción al estudio comparado del comportamiento*, Ediciones Omega, Barcelona, 1974.

HARRIS, CHR., "Psychophysiological Responses to Imagined Infidelity: The Specific Innate Modular View of Jealousy Reconsidered," *Journal of Personality and Social Psychology*, v. 78, n. 6, (2000), pp. 1082-1091.

HARRIS, CHR., "A Review of Sex Differences in Sexual Jealousy, Including Self-Report Data, Psychophysiological Responses, Interpersonal Violence, and Morbid Jealousy," *Journal of Personality and Social Psychology*, v. 7, n. 2, (2003), pp. 102-128.

HARRIS, CHR., "Factors Associated with Jealousy over Real and Imagined Infidelity: An Examination of the Social-Cognitive and Evolutionary Psychology Perspectives," *Psychology of Women Quarterly*, v. 27, (2003), pp. 319-329.

HARRIS, CHR., "Male and Female Jealousy, Still More Similar than Different: Reply to Sagarin (2005)," *Personality and Social Psychology Review*, v. 9, (2005), pp. 76-86.

LEDOUX, J., *The Emotional Brain*, Simon and Schuster, New York, 1996. Spanish translation by Marisa Abdala: *El cerebro emocional*, Editorial Planeta, Barcelona, 2000.

ORTEGA Y GASSET, J., *Meditación de la técnica*, Revista de Occidente, Madrid, 1957. English translation by Helene Weyl with revisions by Edwin Williams: "Thoughts on Technology" in MITCHAM, C. and MACKEY, R. (eds.), *Philosophy and Technology. Readings in the Philosophical Problems of Technology*, The Free Press, New York, 1972, pp. 290-313.

RUSE, M., *Sociobiology: Sense or Nonsense?* Reidel, Dordrecht, 1979. Spanish translation by Arantxa Martín Santos: *Sociobiología*, Ediciones Cátedra, Madrid, 1989.

SANMARTIN, J., *Los Nuevos Redentores*, Anthropos, Barcelona, 1987.

SANMARTIN, J., *La mente de los violentos*, Ariel, Barcelona, 2002.

SANMARTIN, J., *La violencia y sus claves*, Ariel, Barcelona, 2004.

SANMARTIN, J. (ed.), *El laberinto de la violencia*, Ariel, Barcelona, 2004.

SANMARTIN, J., *El terrorista. Cómo es. Cómo se hace,* Ariel, Barcelona, 2005.

WILSON, E., *Sociobiology. The New Synthesis*, Harvard University Press, Cambridge, MA, 1975. Spanish translation by Ramón Navarro: *Sociobiología*. *La Nueva Síntesis*, Ediciones Omega, Barcelona, 1980.

WILSON, E., *On Human Nature*, Harvard University Press, Cambridge, Massachussets, 1978. Spanish translation by Mayo Antonio Sánchez: *Sobre la naturaleza humana*, Fondo de Cultura Económica, México D.F., 1980.

Mathematical Doing and the Philosophies of Mathematics

Javier de Lorenzo

> "... á aucune époque les mathematiciens n'ont été entiérement d'accord sur l'ensemble de leur science que l'on dit celle des verités évidentes, absolues, indiscutables et définitives; qu'ils ont toujours été en controverses sur les parties en formation des mathématiques: toujours ils ont estimé que leur époque était une période de crise..."
>
> Lebesgue, H., *Entretiens de Zürich*, 1938.

1. Split between Mathematical Doing-Philosophical Thought

When G. Leibniz terms as "extravagant mathematicians" those who, in addition to devoting themselves to their *doing*, focus on discussing the labyrinth of the continuum, he is splitting the planes of mathematical doing into two: those who do mathematics and those who seek its principles, its "metaphysics."[1] From this split, and for Leibniz, mathematicians should confine themselves to dealing with rules of calculus and not concern themselves with the nature of what the characters they are handling may refer to, a handling with which, however, he builds the conceptual elements of his work. It is up to the metaphysician to give account of the why and wherefore of these rules, as well as the kind of existence of mathematical objects, to debate whether these are useful fictions, individual substances or something else.

This division has hung over mathematics and has been accepted by most philosophers, above all logical philosophers since the twentieth century, who have not seen or wished to see the ideology underlying Leibniz's self-contradictory position. This split has led to controversy and dispute and has, in general, kept mathematicians away from any public debate of their ideas. What is more, it has led mathematicians to distance themselves somewhat from the literature of philosophers when the latter address the "metaphysics" of *doing*. This literature has gone from dealing with the "fundamentals" to concerning itself with the "philosophy" of mathematics and in most works there is no clear distinction between what may be termed "basic" mathematics and what "advanced" mathematics. These works focus on the natural number, issues related to sets, questions of logic and, at most, some touch on the theme of categories. Mention is made of principles, references, meanings, truth... and nothing, or very little, of what mathematicians feel to be their *doing*, leading them to distance themselves from this kind of writing, if not from the true thought on their doing. Two cases reflect this distancing to which I refer.

In the *Preface* to Nicolas Luzin's 1930 work, *Lecons sur les ensembles analytiques*, H. Lebesgue offers a magnificent portrayal of the dissonance between philosopher and mathematician, the impossibility of a common language and understanding.[2] He depicts

---

[1] Cf. Leibniz, G. W., *Discours de métaphysique*, 1686. Introduction and translation by J. Marías: *Discurso de Metafísica*, Revista de Occidente, Madrid, 1942, paragraph 10.

[2] Cf. Lebesgue, H., "Préface," in Luzin, N., *Leçons sur les ensembles analytiques et leurs applications. Avec une note de W. Sierpinski*, Gauthier-Villars et Cie., Paris, 1930. Reprinted in Chelsea Publishing Company, New York, 1972, pp. VII-XI.

the mathematician's refusal to engage in any dialogue come controversy, preferring to focus solely on his doing, leaving aside any involvement in the argument and confining himself, much to his regret, to his praxis in which he implicitly states his philosophy, his concepts of mathematics. This statement, found in Luzin's book, is thus considered by Lebesgue as much a mathematical as a philosophical treatise. In contrast to the expressive style which asserts itself from an attempted realist objectivism and which reaches its height in formalist structuralism —reflecting a specific concept of what is understood by mathematics, and which appeared in 1930, the year which also saw the publication of van de Waerden's *Modern Algebra*—, Luzin explains why he adopts one definition or another, certain approaches and methods over others, discusses what other mathematicians might do, what he himself does and from where he sets out... This book was in total opposition to the formal style which had begun to assert its presence since Göttingen.

If the above-mentioned case highlights the distancing between mathematician and philosopher, the following shows a certain disregard for those who dogmatise in a particular field of mathematics without fully understanding what they are dogmatising about. A caveat: to judge this case, the issue in question must be understood, since certain philosophers may feel offended and attack the mathematician Dieudonné in defence of philosophers when they should in fact smile at the fatuousness, since one of the creators of Bourbakism quotes Poincaré...

Cited below are paragraphs from a private letter sent by J. Dieudonné to Pierre Dugac, May 12th 1984, published in 2003, concerning the argument between H. Poincaré and B. Russell over the fundamentals of geometry in which Russell was introduced to the French mathematics community by L. Couturat. I highlight the final moral which Dieudonné draws and which would be firmly upheld by many other mathematicians:

> "La polémique Poincaré-Russell est bien instructive; elle montre á l'evidence la *nullité* des propos du prétendu "mathematicien" Russell sur tout ce qui touche aux mathématiques; il devait étre complétement autodidacte en la matiére, car ce qu'il dit montre qu'il ne connaissait apparemment rien de tous les travaux sur les fondements de la géométrie, depuis Cayley jusqu'á l'ëcole italienne en passant par Pasch et Klein (pour Hilbert, il est posible, vu la date de la lettre de Couturat, que son livre n'ait pas encore été publié á cette époque). Je trouve que Poincaré est bien bon de prendre le temps de discuter ce verbiage, et expliquer tout au long le Programmme d'Erlangen (sans le citer, sans doute pour ne pas effaroucher les lecteurs de la *Revue de Métaphysique et de Morale*); c'est d'ailleurs peine perdue car Russell n'y a rien compris, et continue d'employer les mots "vrai" et "faux" á tort et á travers á propos des Mathématiques et de leur rapport au réel (mots que Poincaré avait eu grand soin d'eviter). Il est vraiment dommage que Couturat ait été impressionné á ce point par un tel personage, mai malgré les louables efforts qu'il avait faits pour s'initier aux mathématiques, il est clair que ses connaissances étaient bien insuffisantes, comme le montre la question que Lachelas et lui posent á Poincaré á la fin de sa lettre, oú ils confondent la géométrie projective et la géometrie lobatchevskienne!! Moralité: les philosophes feraient mieux de connaïtre les mathématiques avant de pretender en parler!"[3]

[3] DUGAC, P., *Histoire de l'Analyse*, Vuibert, Paris, 2003, pp. 221-222.

2. The Mathematician's Work, Mistakes, Philosophy

The split highlighted by Leibniz has created an image in which mathematician are seen —in Leibnizian terms— to confine themselves to calculating, to deducing based on previous points which have been clearly established: the initial axioms, the early definitions, with rules of derivation which, if not explained are at least supposed...; where mathematical concepts are perfectly and clearly laid down by those definitions and axioms, and statements or propositions are put together in theories which, if not formalised, are at least axiomatised. This image shows the mathematician to be a mere prover of theorems, one who strengthens the hypothetical-deductivist facet, and at the same time, syntactic formalist nature of mathematics.

This image implies a vision of mathematical praxis which is clearly insufficient for even the mathematical creator Leibniz since, by his own confession, this creation is based on insight and intuition and not on any proof or calculation in which he was often wrong. It is an image which proves inadequate despite its indiscriminate use as an ideological basis, particularly by philosophers who have neither set out nor resolved a mathematical problem, have never offered a mathematical proof, and yet insist on what they feel to be their right: to speak about what they have no experience of, holding forth that the task of the mathematician is simply to "prove" theorems.

As a mathematician, I take completely the opposite stance: I do not see the mathematician as a black box, as the American sausage factory —to use Poincaré's metaphor— into which pigs are fed at one end and out of which emerge the carefully manufactured strings of sausages and legs of pork... In stark contrast, I see the mathematician groping about in his work, at times succeeding at others failing, his intuition proving to be right, coming up with Mediterranean "discoveries" ... led by one concept or another and, in consequence, doing one kind of mathematics or another, criticising or accepting what others do, how they do it, in a constant state of crisis and controversy. His work also symbolises *his* philosophy of mathematics in which he embodies his idea both of what he does and why he does it. To take the words of Gustave Choquet,

"l'activité mathématique ne se réduit pas aux théorèmes pasteurisés qui dorment dans les revues des bibliothèques."[4]

Apart from a few exceptions, it is a somewhat localised philosophy —concerned on occasions with the "metaphysics" of calculus, on others with geometries and space, or arithmetic, or the role of existential proofs...—, without any major desire to establish fundamentals or seek conclusive truths. Many times this philosophy dare not express itself for fear of the outcry from fools, most mathematicians preferring to maintain "public silence" with regard to their thoughts and reveal themselves only in private letters, conversations with colleagues or in works —such as C. F. Gauss, Ch. Hermite, L. Kronecker, N. Luzin, G. Gentzen...— which are mathematics in appearance only.

Mathematics is a praxis which is not alien to thought and is thus a living praxis, not split into two, as Leibniz and the philosophers who follow him would have it, with one part for the mathematician —mathematical doing— and another for the philosopher —unfolding

---

[4] Choquet, G., "Préface", in Lebesgue, H., *Les lendemains de l'intégrale. Lettres à Émile Borel*, Vuibert, Paris, 2004, p. 5.

the philosophy of what the mathematician does. Conceptions held of mathematics have thus varied, have been subject to change, where the focal point has maybe centred on the search for precision in the concepts which are being dealt with at each moment. These concepts are not stated once and for always, packaged and boxed. Mathematical doing focuses to a great extent on the search for precision of concepts, on the changes which this leads to in these concepts and the changes in the properties attributed to it, and thereby changes in theories which emerge. The mathematician is not confined to proving theorems from previously established perennial principles, but rather creates concepts, solves problems, and by his action, catches a glimpse of what may subsequently be called early principles.

As P. S. Laplace wrote in January 1792, when encouraging Lacroix in his efforts to construct a treatise which might bring together individual propositions and highlight the relation between the thoughts with which he might thus achieve general methods over and above individual results:

> "Le rapprochement des méthodes que vous comptez faire sert á les éclairer mutuellement, et ce qu'elles ont de commun renferme le plus souvent leur vraie métaphysique; voilá pourquoi cette métaphysique est presque toujours la derniére chose que l'on découvre. Le génie arrive par instinct aux résultats; ce n'est qu'en réfléchissant sur la route que lui et d'autres ont suivie qu'il parvient á généraliser les méthodes et á en decouvrir la métaphysique."[5]

I shall henceforth aim to put myself in the position of those who create or build mathematics and not in that of those who disseminate, use, teach or apply it for other purposes or for their "own personal philosophy," with the dangers this choice implies.

Mathematics is constructed by the mathematician, who is born at a specific period in history and in specific circumstances, thereby inheriting previous theory and subject matter, certain types of problems, approaches which entail varying mathematical styles and, of course is influenced by trends or preferences as well as professional constraints.

The mathematician acquires products through texts and courses which have their own structure just like the building in which we find ourselves at this moment. These texts, as such, are neither true nor false, but simply exist, like the building in which we stand. This has been constructed within a subject area —Analysis, Geometry, Topology, Algebra, Number Theory...—, with the materials available at the time —Gauss's is not the same as Bourbaki's— and with a specific goal: whether didactic, for dissemination, as a creative essay, a paper written and published for a "productivity bonus" in salary as is the case nowadays, for resolving problems, for application in other disciplines... The text denotes the expressive style of the moment in which it is written by the author as well as reflecting the time and school to which it belongs, although it cannot reflect the whole mathematical doing of the moment.

2.1. Non-Accumulative Doing

Not all the doing contained in each of these texts and courses is kept through time. A large part of what mathematicians do is lost, not only in the past but also at the present.

---

5 Laplace is quoted in Dugac, P., *Histoire de l'Analyse*, p. 69.

There are thousands of propositions in PhDs, for instance, which remain and will always remain locked away. Propositions, problems, and questions stored in libraries and magazine libraries, and nowadays on computers, which may one day see the light of day when some future historian dusts them off.

There are objective reasons for this. A particular kind of mathematics may form part of a more general theory and thus lose its identity. It may also prove to be too out-dated, as if it had disappeared even though it may still be used in certain circles. Such is Rey Pastor's lament when faced with von Staudt's synthetic projective geometry, more deeply studied in Spain than in Germany itself, and which led him to assert that his own work on geometry had, in some way, become antiquated, a veritable relic compared to mathematics which, by his own confession, took a different course.

Here we may mention the question posed in 1972 by Bernkopf when studying the life and work of Halphen,[6] asserting, "the sum and quality of Halphen's work is outstanding" leading him to wonder:

> "Pourquoi, alors, son nom est-il si peu connu? La reponse se trouve partiellement dans le fait qu'une partie de son oeuvre, la théorie des invariants différentiels, est maintenant seulement un cas particulier de la théorie plus genérale des groupes de Lie et qu'ainsi elle a perdu son identité. Mais une partie de la reponse se rattache á une question plus importante: pourquoi tant de mathématiques mëme d'une passé récent sont-elles perdues? Dans le cas de Halphen, il a travaillé sur la géometrié analytique et différentielle, un sujet si demodé aujourd'hui comme s'il était presque disparu."[7]

It is not just the work —and the figure— of Rey Pastor, of Halphen but that of so many others, the very memory of whom has been lost or who, at most, are linked to a name, a theorem or property which, in many cases, they may not even have formulated... Results and products which lie buried in works which, when an attempt is made to rescue them from oblivion, are in most cases illegible to the modern-day reader.

A third reason may hold the key. Mathematical concepts change and their origin as well as part of their evolution become relics or fossils, suitable for the historian or for those introductions to the subject which certain educationalists see as the most natural genetic approach, rather than going to directly to the basic concepts of the moment.

Mathematical doing is not merely accumulative. Materials, style, problems, and disciplines change, as do the needs of mathematicians and the society in which they are immersed. It is mainly concepts that vary, are transformed, implying that mathematical doing is not exact, if by exactness one understands the total perfection of concepts, the correctness of proofs.

2.2. The Mathematician's Error

Emerging from these transformations one may speak of a lack of exactness at a given moment, of the mistakes made by all mathematicians in their doing, something which history books on the subject try to gloss over. In traditional philosophy of mathematics the role of error in mathematical work has generally been neglected. Yet, error is vital, and

6 Cf. Bernkopf, M., "Halphen, Georges Henri," *Dictionary of Scientific Biography*, vol. VI, Scribner's and Sons, New York, 1972, pp. 75-76.

7 Quoted in Dugac, P., *Histoire de l'Analyse*, p. 243.

is a key driving force in mathematical praxis. An awareness of error leads to it bearing fruit through the creation of an ensuing methodology: the *counterexample* which serves to destroy, while at the same time, force the transformation of what has been destroyed.

There are different types of error. An author may clearly suffer a "distraction" and adopt a definition omitting one condition therein —for example, Lebesgue takes Borel's theorem and overlooks the term denumerable; a distraction which leads to the theorem of Borel-Lebesgue. A mistake might be made when proving a theorem —by not including a possible condition which may be more or less restrictive— and, when attempting to correct it, this error in proof may bear magnificent consequences in the shape of new ideas and new concepts.

Without resorting to the hundreds of erroneous proofs of the fundamental theorem of algebra, one illustrative case may be cited from Lebesgue's report *On analytically representable functions*, in which he states and proves the theorem "the projection of an intersection of measurable sets B is always a measurable set B." In the words of Lebesgue "the demonstration was short and simple, but wrong." When attempting a new proof, Luzin reaches the conclusion that the theorem is wrong because the projection does not always give a measurable set, but a new type of set whose complementary is, in general, neither measurable nor analytical... Luzin inspires a new field, analytical functions and sets, extending the field of Analysis known up until then. "Fruitful error" as Lebesgue calls it in his *Preface to* Luzin's book.[8]

Continuing with Lebesgue rather than other mathematicians such as B. Riemann, K. Weiertrass, H. Poincaré... another productive error is found in his affirming that the solutions to G. Dirichlet's problem had the same proprieties of continuity in the plane as in space. Spurred on by certain doubts, Lebesgue constructs, by way of a counterexample, a surface in space with one particular feature today known as "the Lebesgue thorn." This peculiarity led to the concept of the irregular point, crucial to the potential theory.

There are other kinds of error based on the fact that, as is often remarked upon in history books, the author has neither the suitable methods nor technical means available to perform the correct proof; means only subsequently accessible and with which the author's inevitable mistakes in his early attempts at proof can be revealed and corrected. This has often been pointed to in certain of Poincaré's works,[9] particularly his fourth report on Fuschian functions, although this attribution is incorrect as it was his contemporaries who failed to see the conceptual content in Poincaré's work.

Together with the examples cited there is the problem arising from the fact that the concept in question is not clear, the most common case. Take for example the difference between convergence and standard convergence, the notion of function, the distinction between differentiability and continuity, differences which, prior to emerging, lead to their handling which only through those differences is considered erroneous. In certain instances these cannot be deemed errors in the strict sense —the differentiability-continuity question, for example— but rather show that the concepts are not given once and for all but vary and lead to a situation where, at each moment, the concepts are handled with the utmost rigour by mathematicians.

---

[8] Cf. LEBESGUE, H., "Préface," in LUZIN, N., *Leçons sur les ensembles analytiques*, p. VII.

[9] Cf. POINCARÉ, H., "Sur les groupes des equations lineaires," *Acta Mathematica*, v. 4, (1884), pp. 201-312. Translated into English by J. Stillwell: POINCARÉ, H., *Papers on Fuchsian Functions*, Springer, New York, 1985.

The mathematician is immersed in a doing which, as such, is undertaken by others like him, those who are qualified and who have emerged at a given time and place. They find their materials, styles, and schools. Yet, the mathematician-creator must go beyond this, and not confine himself to the mere application of rules and methods. By going beyond them, he becomes aware that he is basing his doing on an idea, a philosophy, *his* philosophy, of what he is engaged in.

3. Some Exemplary Schemes

By rejecting eternalist statism as something unbecoming of any type of human praxis, history can illustrate the conviction, the belief that the mathematician, in his constructive doing, is led by a conception, what we may call his philosophy. He is not the extravagant mathematician, as Leibniz would have him. The key to this doing is not based, however, on what may be proved but, fundamentally, on what may be conceived, with a clear awareness that, at each point, mathematicians are discussing among themselves what they are doing and how they are doing it.

History may provide us with four schematic examples both in terms of intrinsic development as well as in the number of cases chosen. These should not give the impression that the mathematicians mentioned confined themselves exclusively to "pure" mathematics, although no reference is given here to their contributions in other fields, such as physics or probability theory, in which the mathematicians cited intervene, in some cases decisively.

3.1. Lagrange and his Concept of Function

In 1797 J. L. Lagrange published *Théorie des fonctions analytiques.*[10] This was the fruit not only of the course in the Polytechnic but also of his reflections to formulate a general theory on functions in which the notion of the infinite is dismissed together with any consideration of the infinitely small or indivisible in mathematical doing. This work, and from the outset, a thesis takes an ontological standpoint on the objects used by the mathematician as well as a methodological position as to how they should be handled.

What is at stake is the concept of function that Lagrange restricts to "analytical function." For Lagrange all functions of a variable x are made up of the "simple analytical functions of a single variable" such as x^n, ax, log x, sin x and cos x by the four arithmetical operations, namely, addition, subtraction, multiplication and division or, in other words, by algebraic composition or "are given in general by equations in which functions of these same types are involved."

Differential calculus is simply reduced to "the algebraic analysis of finite quantities" since all analytical functions may be developed in series. Lagrange points out that if instead of x we put x + i, with i any indeterminate, we obtain f(x + i), and by the theory of series, it may be developed as $fx + pi + qi^2 + ri^3 + \ldots$ where the quantities p, q, r, ... , coefficients of the powers of i, will be new functions of x, derived from the primitive function fx and independent of the quantity i.

---

[10] Cf. Lagrange, J. L., *Théorie des fonctions analytiques, contenant les principes du calcul différentiel, dégagés, de toute considération d'infiniment petits ou d'evanouissans, de limites ou de fluxions, et réduits a l'analyse algébrique des quantités finies*, l'Imprimerie de la République, Paris, 1797.

This way of deducing one function given other functions, which are derived functions depending only on the primitive function, is what Lagrange considers to be "the real goal" of new calculus, differential calculus. Lagrange breaks new ground in that these derivatives are generated as functions by their analytical expression from the canonical analytical expression through Taylor series development of the primitive function. As a consequence, the functions appear as the result of extremely basic algebraic calculus, arithmetical in essence; what has come to be called Lagrangian formalism.

Analysis cannot be based on such a dark concept as continuity, as appears when speaking of fluxions for instance. Yet, nor can it be grounded on the even less rigorous concept at the time —despite the efforts of d'Alembert— of the *limit* of reasons. It is in the concept of series on which Lagrange bases his analysis and, from this notion, attempts to establish the concepts of derivate, integral...

For Lagrange, this approach endows Analysis with the necessary rigour of earlier proofs, aiming to reduce the infinite to the finite through the purely algebraic —or rather arithmetic— use of whole series.

Lagrange's work is guided by an ontological stance with regard to the infinite both in the large and the infinitely small. It is not "blind," restricted to following rules of calculus. Here he comes into conflict with other mathematicians who hold other conceptions. He was involved in discussions with Leonhard Euler, Daniel Bernoulli, Jean d'Alembert on the concept of function, its possible expression through a trigonometric series with infinite terms given at the moment, the role which cancelling out the second partial derivative at a point might have, d'Alembert's insistence, in opposition to Euler and D. Bernoulli, on eliminating any reference to geometric or physical representation in the case of arbitrary solutions of the equation corresponding to vibrating strings, fully accepted in his work in physics.

D'Alembert only admits "continuous" solutions, in other words, analytical functions defined by a single law as opposed to Euler who accepts these analytical functions but also those which are "mechanical" or "discontinuous" —those traced freely by hand— or "mixed" comprising different interlaced curves which may be expressed in sections "by a law or an equation," perforce different for each section. For d'Alembert the *metaphysical reason* preventing the existence of solutions in certain cases is, precisely, the jump in the curve at one of its points, a jump given by the nullity of the second derivative.

One of the mathematician's intrinsic goals is thus revealed: to provide the "true metaphysics of differential calculus," not to restrict himself to problem solving, proving theorems. D'Alembert interprets metaphysics as illustrating "the general principles on which a science is based," these principles not being what is today understood by propositions or axioms but rather the basic conceptual or key elements for developing Analysis. From the conceptual, the *limit* concept is the basis of that "true metaphysics" for d'Alembert; for Lagrange it is that every function may be developed in series.

Proofs of theorems or principles that are felt to be fundamental remain, from another point of view, open to question. The considerations and criteria for series convergence do not appear as essential in Lagrange's work, his supposed rigour, pure entelechy from that subsequent point of view. Even his conviction that a derivable function —and therefore a continuous one, developed in series, reduced to the algebraic analysis of finite quantities—, does not require the actual infinite. This is explained by in 1821 by Cauchy when attacking Lagrange's formalist programme:

> "Substituting the functions for series, implicitly supposes that a function is completely characterised by development of an infinite number of terms, at least in so far as these terms have finite value."[11]

It is Lagrange's metaphysics that remains open to question, based on convictions or acts of faith which are not proved and thus, deep down, self-contradictory, as pointed out by Abel in 1826 in a letter to B. Holmboe:

> "Le théoreme de Taylor, la base de toutes les mathématiques supérieures, est tout aussi mal fondé. Je n'en ai trouvé qu'une démonstration rigoureuse, et elle est de Cauchy dans son *Résumé des lecons sur le calcul infinitésimal*."[12]

A. L. Cauchy, after this rigorous proof offered by N. H. Abel, goes on to show that if the Taylor series of a function f converges at a point x, it does not necessarily have to be equal to f(x). To do this he uses the counterexample method, as Abel had done with Cauchy himself. The counterexample given is the function f defined by $f(x) = e^{-1/x2}$ if $x \neq 0$ and $f(x) = 0$ if $x = 0$. A counterexample that cancels out Lagrange's conviction that functions f(x), f'(x), f''(x), ... "cannot become null at the same time by the supposition x = a, as certain geometricians would seem to suppose." In this case $f^n(0)$ is null for the whole of n as a result of which Taylor's series is likewise null and yet, the function f(x) is not.

What is now clearly at stake is not whether, by exclusion, a function may be developed in series or not but rather the concept of function itself. By way of a conceptual clash, Cauchy's counterexample, as P. Dugac points out, "opened the way for the theory of quasi-analytical functions, one of the favoured topics of analysts prior to the Second World War."[13] Yet, we are forced to admit that Lagrange's formalist-algebraic approach, contested by Cauchy and his followers, is reborn with the work of Ch. Méray —in the related history books with those of Weiertrass— when he establishes an arithmetic theory of irrational numbers in 1869 that is extended in his 1872 and 1894 books on infinitesimal analysis. Méray relies on the certainty that

> "all analytical functions have the common property conjectured by Lagrange that they *may always be developed in whole series, in other words by Taylor's formula*. (...) replacing this general property with continuity, monogeneity, etc., I choose it as the sole base of all reasoning."[14]

Like Cauchy and Méray, Lagrange has not developed his work blindly but led by the "true metaphysics" of what he was doing, seeking key concepts to build mathematical Analysis and endow it with the necessary precision which, in his opinion, it lacked. This is clearly manifested in the full title of his book: *Containing the Principles of the Differential Calculus Disengaged from All Consideration of Infinitesimals, Vanishing Limits or Fluxions and Reduced to the Algebraic Analysis of Finite Quantities***.**

---

[11] ABEL, N. H., *Correspondance. Mémorial*, Kristiania, 1902, p. 17. Cited in DUGAC, P., *Histoire de l'Analyse*, p. 103.

[12] Quoted in DUGAC, P., *Histoire de l'Analyse*, p. 103.

[13] DUGAC, P., *Histoire de l'Analyse*, p. 77.

[14] MÉRAY, CH., *Lecons nouvelles sur l'analyse infinitesimales et ses applications géométriques,* Gauthiers Villars, Paris, 1894, pp. xv-xvi (Author's emphasis).

3.2. The "Italian School" of Projective or Algebraic Geometry

The geometry of greater than three-dimensional space heralded one of the most significant achievements of the Italian School, as did the geometrisation of number theory. Luigi Cremona (1830-1903) is felt to be the creator of this School with his development of the geometric (synthetic) theory of plane curves and his analysis of birational quadratic transformations of the plane through what are today known as Cremona transformations, linked to cone surfaces. After Cremona, in the field of projective geometry, the door to algebraic geometry, appears the figure of Corrado Segre (1863-1924) followed by G. Castelnuovo (1865-1952), F. Enriques (1871-1945), F. Severi (1879-1961), ...

According to José Manuel Aroca,[15] Segre systematically uses projective generation arguments and projection from hyperspaces to build new three-dimensional figures and analyse their properties, innovative for the period since around the 1880's spaces with a dimension greater than three lacked any entity for some mathematicians.

The mathematicians in this school, whose work was followed not only in Italy but by mathematicians like J. L. Coolidge and S. Lefschetz and studied in Spain up to the 1960's —such as Eugenio Bertini's *Complements of Projective Geometry* published in 1927, translated by Germán Ancochea in 1951 and studied as a "complement" in the Faculty of Mathematics in Madrid—[16] based their work on a very clear conception of what their doing was. For Severi, points, straights, planes, etc. of an n dimension space were

> True geometric entities and not mere attributes of analytical entities. The lineal space of n dimensions really exists, and is not reduced to the shadow of linguistic fiction.[17]

From this standpoint, algebraic reductionism —mere linguistic fiction— lacked any sense and with it geometric algebra. The prominence of the geometric over the algebraic and the analytical is based on intuition, an intuition which reaches the extreme reflected in André Weil's anecdote about Enriques who, when addressing a student, remarked:

> "You can't see it? What do you mean, you can't see it? I can see it as clearly as I can see that dog over there."[18]

Intuition is the key, yet not naïve intuition because hyperspace cannot be perceived, but rather "theoretical" intuition, which is not merely a linguistic device through which geometric language is handled to guide what is essentially considered to be analytical or algebraic study. Based on geometric intuition, Segre achieves remarkable progress in areas which are today termed as lineal algebra although it was to be G. Peano —diametrically opposed to this geometric school and creator in turn of another Italian school with a formal axiomatic tendency— who would establish the first complete definition of vectorial lineal (real) space, in its formal structural, axiomatic sense.

Segre, in 1891, in the first issue of *Revista di matemática* published by Peano, outlines "his programme" in which, as a starting point, he demands that

---

[15] Cf. Aroca, J. M., "Zariski y la Escuela italiana," in: *Historia de la Matemática en el s. XX*, Real Academia de Ciencias Exactas, Físicas y Naturales, Madrid, 1998, pp. 171-193.

[16] Cf. Bertini, E., *Complementos de Geometría Proyectiva*, Aguilar, Madrid, 1951.

[17] Cf. Aroca, J. M., "Zariski y la Escuela italiana," in: *Historia de la Matemática en el s. XX*, p. 174.

[18] Aroca, J. M., "Zariski y la Escuela italiana", p. 171.

Science needs to be given as free a hand as possible.[19]

The purity of the method takes on a secondary role even if we must sacrifice accuracy, a major sacrifice when dealing with mathematics. Unlimited space may be introduced without any link as regards the number of dimensions and without any restriction for potential applications, in relation to which he adds:

> "It has been noted... that for instance the general projective geometry that we built in this way, is not anything other than the algebra of linear transformations. This is only a difference not a fault. Provided that we do mathematics!"[20]

Geometric intuition, without algebraic reductionism, reaches the level of method above any "rigorous" proof, although the latter may subsequently be required. This position was immediately criticised by Peano after Segre's essay, and it is worth remembering that Segre and Peano worked in the same faculty in Turin and were about the same age. He was joined in this criticism by Castelnuovo, Enriques, Severi who came out in defence of abstract but geometrically intuitive mathematics as opposed to the formal rigour demanded by Peano. If taken to the extreme, this formal rigour may lead to a "sclerosis" of the mathematician's creative thinking. The Intuition-Rigour conflict reaches the point where, as Zariski relates, if anyone pointed out to Enriques that something was missing in a mathematical proof, the latter would answer:

> "Oh, come on! That's just *dubbio critico* (critical scruples)"

Zariski is also reported to have said

> "Theorems are aristocratic, demonstrations plebeian."[21]

Together with geometric intuition, the mathematician must use analogy and experimentation as basic tools. As Severi writes in 1937, after the works of E. Nöther, van der Waerden...

> "Analogies are often precious; but in many other cases they are prisons where for lack of courage the spirit remains fettered. In any branch of science, however, *the most useful ideas are those that born orphans later find adoptive parents*.
>
> New mathematical constructions do not actually proceed exclusively from logical deduction, but rather from a sort of experimental procedure making room for tests and inductions —these constitute the real ability necessary for exploration and construction. And thus is it natural that theory development entails some modifications of, and adjustments to, initial conceptions and the language expressing these conceptions; and that one can reach a definite systematisation only by successive approximations."[22]

All the mathematicians of this school were fully conscious of this position, as were certain others, more distant, who shared this opinion such as Félix Klein who in 1893

---

[19] Cf. Brigaglia, A., "The Creation and Persistence of Natural Schools: The Case of Italian Algebraic Geometry," in Botazzini, U. and Dahan-Dalmédico, D. A. (eds.), *Changing Images in Mathematics*, Routledge, London, 2001, p. 193.

[20] Brigaglia, A., "The Creation and Persistence of Natural Schools: The Case of Italian Algebraic Geometry," p. 193.

[21] Both quotations from Aroca, J. M., "Zariski y la Escuela italiana", p. 171.

[22] Brigaglia, A., "The Creation and Persistence of Natural Schools: The Case of Italian Algebraic Geometry", p. 196; emphasis added.

spoke of the *crisis* that mathematics was suffering due to the conflict between intuition and rigour.[23] This position led to work that was far removed from the formal Peanian approach and subsequent algebraic structuralism, which came to the fore after the work of Emmy Nöther in the field of algebraic geometry, where geometry is subsumed into algebra.

The position and work of the mathematicians in this Italian school was to become antiquated, not merely due to the triumph of the algebraic structural approach but because this conception marked the boundaries in the use of purely geometric methods and in the use of intuition which proved hard to come to terms with. Zariski —who despite working with the Italian mathematicians was not "one of them"— is considered to have finally evidenced these limits once and for all. In the words of Aroca, Zariski is considered

> "The last of the great classical Italian geometricians and the first of the great modern geometricians."[24]

Aroca's words are significant as they juxtapose "classical" and "modern"... Zariski is the bridge by which the contribution of Italian algebraic geometry came to form a key part of algebraic geometry, at the cost of undergoing a radical transformation, so radical that it became unrecognisable. In his 1990 biography, Parikh points to the words of Zariski:

> "I wouldn't underestimate the influence of algebra, but I wouldn't exaggerate the influence of Emmy Nöther. I'm a very faithful man… also in my mathematical tastes. I was always interested in the algebra which throws light on geometry and I never did develop a sense for pure algebra. Never. I'm not mentally made for purely formal algebra, formal mathematics. I have too much contact with real life, and that's geometry. Geometry is the real life."[25]

In *Commutative Algebra*, published in 1958 and written in collaboration with Pierre Samuel, the Preface begins with a statement which picks up, I am not sure whether consciously or not, on the words of Severi previously highlighted:

> "This book is the child of an unborn parent."[26]

As in the case of Lagrange, among the mathematicians of the Italian School of projective-algebraic geometry there is a belief that drives and guides their praxis in the clear knowledge that their approach conflicts with other currents.

3.3. French School or School of Paris

At the same time as the Italian school mentioned earlier, and in clear confrontation with another Italian school, namely Peano's, there emerged in Paris a group of leading mathematicians who worked with set theory, paving the way for a mathematical world which was radically different to the two schools previously cited despite being contemporary. The world was also different, reflected by the tremendous concern G. Cantor seemed to show

---

23 Klein's paper was delivered on 2 September 1893 in the Sixth Colloquium Evaston. It was published in *The Evaston Colloquium Lectures on Mathematics*, MacMillan, N. York, 1911. Cf. Ewald, W., *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, Vol. 2, Clarendon Press, Oxford, 1996, p. 959.

24 Aroca, J. M., "Zariski y la Escuela italiana", p. 171.

25 Brigaglia, A., "The Creation and Persistence of Natural Schools: The Case of Italian Algebraic Geometry", p. 199.

26 Cf. Zariski, O. and Samuel, P., *Conmutative Algebra. Vol. I*, van Nostrand, N. York, 1958, p. v.

at the time and which was restricted to questions of the transfinite and the problem of the hypothesis of the continuum. French mathematicians adopted set theory as the basis of their constructions on Analysis. In particular, E. Borel (1871-1956), René-Louis Baire (1874-1932), H. Lebesgue (1875-1941), M. Fréchet (1878-1956), all of whom were Normalians. They developed both Analysis and the theory of functions of real variable, Borelian sets with their link in Lebesgue's measure and integral, metrics and compact spaces...

Versed in the work of Cantor and his methods such as the diagonal, they accepted those results but with a great deal of criticism founded on the change to transfinite powers without remaining in the numerable. They convert the actual infinite into a problem, and what it implies in existential terms, transferring the result of that criticism to their mathematical praxis.

In 1898, Borel published *Lecons sur la théorie des fonctions*, subtitled *Principles of the theory of sets in the light of applications to the theory of functions*. It is not the mere application of set theory, but is also a critique of some of its aspects. With great clarity, it poses the problem of the intelligibility of Global Doing, pitfalls inherent in mathematical conceptualisation reflected therein. Global Doing points to "a function of any variable" without further specification, unviable for a constructivist mathematician. For Borel, mention of "any" function requires clarification:

> "*Fonction quelconque* on entende une fonction qui puisse étre *effectivement* définie, c'est-á-dire telle qu'on puisse, par un nombre limité d'opérations, calculer, avec une approximation donné, sa valeur pour une valeur donné de la variable."[27]

Jules Tannery, René-Louis Baire, E. Borel, H. Lebesgue pose the problem of the existential question of mathematical objects, and with it whether adding new modes of reasoning is feasible. They adopt a critical attitude towards the existence of objects which may not be constructed, linking construction to the term definition, where it is impossible to refer to "any" function if there is no effective means of defining it, constructing it or calculating the corresponding value for each argument. It is not a question of speculating on independent doing, but of their own praxis, their work as mathematicians. The question is whether or not it is possible to reason "in the vacuum", to reason with "any" functions, with objects it is not known how to specify, how to obtain.

In a letter to D. Hilbert in 1904, E. Zermelo shows that the continuum may be well ordered, to prove which he explicitly uses the Axiom of Choice. Zermelo's proof heightened the debate between mathematicians and the following issue of *Mathematische Annalen*[28] saw the appearance of the critical comments of S. N. Bernstein, A. Schoenflies, J. König, P. E. B. Jourdain, E. Borel. In his Note, the latter rejects Zermelo's proof based on the arbitrary transfinite choice of an element in each of the infinite disjunctive sets and concludes by saying "such reasonings are outside the realm of mathematics."

Borel enlarged his book of 1898. He was able to involve other French mathematicians in an exchange of letters published in 1905, which Borel collected as Notes for subsequent

---

[27] BOREL, E., *Leçons sur la Théorie des Fonctions*, Gauthiers-Villars. Paris, 4th ed., 1950, p. 118 (Author's underlining).

[28] Cf. ZERMELO, E., "Beweis, daß jede Menge wohlgeordnet werden kann," *Mathematische Annalen*, v. 59, (1904), pp. 514-516.

editions of the previously cited book. Two sides come into play: Hadamard defends both the Axiom of Choice as well as Zermelo's proof; and Baire, Borel as well as Lebesgue reject both questions.

The problem centres on scientific conceptualisation, and how it may be established. J. Hadamard resorts to J. Tannery's distinction between correspondence that can be *defined* and correspondence that can be *described*. In the Axiom of Choice, defined existence is held, yet cannot be described since there is no way to effectively execute the operation in question. The individual elements that the function of choice chooses in each of the sets of a family of the power set of the set given cannot be characterised or shown. Hadamard accepts this type of non-constructive reasoning. Counterattacking, he points out that in his work Borel "Tu emploies des correspondances dont tu constates *l'existence* sans pouvoir cependant les *décrire*."[29]

The Axiom of Choice appears to be *the* model of a non-constructivist existential approach and is at the heart of these letters because it is this which clearly poses the question of what is understood by the definition of mathematical object and, therefore, whether it is feasible to reason something which can be defined but not described. Borel's position is clear:

> "I do not understand the point of view of analysts who think they can reason on a specific yet not defined individual; there is a contradiction in the terms in which I have often insisted."[30]

In answer to Hadamard, Lebesgue attempts to delimit the matter under debate for which he poses the question: "*Peut-on démontrer l'existence d'un etre mathématique sans le definir?*"[31] But also points out that mere definition does not imply the existence of what is defined. He admits that the word "existence" may be used in many senses and attempts to find a common definition for the term. He proposes linking it to "la loi qui définit les elements choisis"[32] and the fact of this law is to him essential, whether it be a numerable or non-numerable infinity.

Lebesgue picks up on a distinction he established in 1904 in his *Lecons sur l'intégratión et la recherche des fonctions primitives*. Here he distinguishes between a *descriptive* definition in which the principal characteristics of the concept to be defined are established, and is the axiomatic or implicit definition, and the *constructive* definition about which he affirms that

Constructive definitions outline those operations which are necessary to obtain the entity to be defined.

Constructions are mathematical operations which are perfectly defined and this definition is in general given by a law which enables the required value to be obtained.

The term "definition" clearly has its nuances for each of those involved in the controversy. It implies describing or naming a property which is characteristic of what

29 Borel, E., *Leçons sur la Théorie des Fonctions*, 4th edition, p. 150.

30 Borel, E., *Leçons sur la Théorie des Fonctions*, p. 154.

31 *Leçons sur la Théorie des Fonctions*, 4th edition, p. 153.

32 Borel, E., *Leçons sur la Théorie des Fonctions*, p. 155.

is defined, but also providing a means of executing what is defined effectively and in a manner that may be calculated. For Lebesgue, the characteristic of what may be named is sufficient; Borel demands that it be calculable and in finite time, in addition. In both cases the demand for a constructive law in the previous sense of Lebesgue seems clear.

For his part, Hadamard emerges as an idealist and attempts to separate two aspects which for him have merged: the subjective and the objective. In a second letter, he asks that the completely subjective question be left to one side.

> Pouvons-nous ordonner?

which is basic to Baire, Borel and Lebesgue for the purely objective question

> "L'ordination est-elle posible?"[33]

It is the same difference which is posed with regard to the non-contradiction position of Hilbert and of Zermelo who asserts that he has proved that "we do not perceive any contradiction." These are psychological elements that bring properties of our brains into play. An idealist must distinguish between what exists in itself and our subjective capacity to grasp or describe such existence, two levels which for the constructivists cannot be separated: an object does not exist in itself but as the result of the mathematician's construction. For Hadamard the objective nature of the mathematical object must be maintained, regardless of whether we know it or not and how. An axiom such as the Axiom of Choice is thus feasible as it is by no means clear that we can in fact express the law that will govern the choice of the element e in each set E. Hadamard would ask: Need a law be explicitly formulated in order to exist? As he is aware that for Lebesgue and Borel this is a rhetorical question, Hadamard emphatically exclaims

> "Qu'il nous soit imposible, au moins actuellement, de *nommer* un élément de cet ensemble, j'ens conviens. C'est lá la question pour vous; ce ne l'est pas pour moi."

Two positions, "two conceptions of Mathematics, two mentalities" which are opposed to one another as recognised by Hadamard,[34] which do not limit themselves to dialectic exercises and where set antinomies or paradoxes are secondary considerations in contrast to what the history books would have us believe. They do agree, however, that a notion such as the set of all sets is antinomical.

Borel attempted to go on with the finitism which he here proclaimed, and commenced a constructive analysis founded on the concepts of real computable numbers, the succession of computable numbers, computable function. This is an analysis where the calculable or computable functions are necessarily continuous and where a principle such as the excluded middle must be used with certain caution since during the decimal development of two calculable functions the n first terms may be equal and there might not exist the possibility of deciding whether one is greater or less than the other. A theory formulated in *Computable Analysis* where computable numbers are shown to form a closed body which contains algebraic numbers as well as other significant ones such as π or e, and the cardinal of the body is $\aleph_0$. Borel's demands are met: no more elements are required to do the mathematics.

---

[33] *Leçons sur la Théorie des Fonctions*, 4th edition, p. 156.

[34] Cf. Borel, E., *Leçons sur la Théorie des Fonctions*, 4th edition, p. 156.

For his part, Lebesgue tries to sidestep the Axiom of Choice and focuses on the use of inductive reasoning, and on the transfinite. As Medvedev points out, to avoid any reasoning that the Axiom of Choice might involve, Lebesgue

> "developed a method of constructing point sets based on the use of the set of transfinite numbers of second kind; moreover the transfinite sequences of real numbers occurring in the set under consideration are defined not simultaneously as an actual set, but successively, so that the definition of each number of the sequence depends essentially on the definition of the preceding numbers. In this way each of the sets considered by Lebesgue is in certain sense constructive."[35]

By constructive definition Medvedev is here referring to that which is not implied by the Axiom of Choice or all transfinite numbers of second kind, considered as if a set were actually given.

In the light of criticism and controversy among mathematicians, Borel as well as Lebesgue strive to develop a theory which is concurrent with their ideas, which they feel to be what mathematical doing should concern itself with.

3.4. Current Situation

After the Second World War, a formalist approach based on axiomatising structuralism came to the fore in the academic, professional and teaching world, in which disciplines restricted themselves to developing what, with Dieudonné, became known as *Bourbakist choice*: algebra, topology and number theory, where scant attention was paid to classical analysis beyond the mere fundamentals. Areas linked to differential equations, algebraic topology, specific geometry, numerical and combinatory methods, probability and statistics were left to one side... as were possible links between mathematical doing and other disciplines particularly Physics. Mathematics withdrew into an ivory tower.

The appearance of this new approach was somewhat surprising as the Second World War had had a major effect on the work of mathematicians and had led many, including those exiled from the Göttingen School to the USA, to work on what was considered "applied mathematics," far removed from the aesthetic purism which led to the formalist structuralist approach. However, a pure approach to mathematics came to prominence in academic circles in the 1950s, a movement which was not only French, prompted by the Bourbakists, but one in which North America was to play a leading role.

This movement sprang up with Hilbert and was consolidated in the early Bourbakists after 1935 with the strictly ideological aim of achieving a definitive and final *doing*. This line was to establish formal structuralism where proof prevails over the content of pure thought. Mathematics appears as a unified and perfectly rigorous whole, for all eternity, based on set theory and the notion of structure. The mathematician works inside closed systems, delimited by initial principles and formal rules that are clearly derived and formulated from which theorem after theorem is obtained. The initial principles, through the implicit definition or axiomatic approach, yield the precise ultimate definitions of the mathematical concepts. This is the materialisation of "Leibniz's dream" I referred to at the beginning.

---

[35] MEDVEDEV, F. A., *Scenes from the History of Real Functions*, Birkhauser, Boston, 1991, p. 162.

This clearly propositional approach is reflected in line with the so-called "linguistic shift" in other fields and considers Mathematics as the logical syntax of scientific languages. There is a prevalence of the formal-deductive, which strengthens the context of justification —the deductive— as opposed to discovery —the creative-constructive— with the *surprise* of the "unreasonable" effectiveness of mathematics for a knowledge of physis. The individual artist's quasi-aesthetic appearance predominates, his creation only being understood by five other creators in the world, if indeed that many.

Occasionally, the mathematician-creator, in his role as artist and oblivious to the call of any other discipline, creates or resolves a problem. Serge Lang, a radical bourbakist in a course in 1982 published in 1984 under the meaningful title of *The Beauty of Doing Mathematics*, writes

> "Every now and again a mathematician solves a problem, which in turn poses others as appealing as the first. These problems are often quite complex and are only understood by experts in the field."[36]

The mathematician Lang refers to, the one who solves a problem, does mathematics but, should there be any doubt, Lang goes on to clarify that

> "By 'mathematics,' I understand pure mathematics."[37]

The Leibnizian *dictum* of an eternalist conception as it has been called is thus fulfilled, one, which not even Hilbert himself, put into practice in his work as a mathematician. This is clearly another philosophy of mathematical doing, with all that it implies and which has prevailed in certain academic circles and has been adopted by most philosophers as I pointed out at the beginning, with the subsequent distancing on the part of the creative mathematicians.

In contrast to the formalist-syntactic image and principally over the final quarter of the twentieth century, there is a reaction in which we still find ourselves immersed, convinced that there is another, less dogmatic, less ideological, way to do mathematics, one which follows more the line of Poincaré and Weyl than Hilbert.

Recalling Weyl's metaphor in 1939, "In these days the angel of topology and the devil of abstract algebra fight for the soul of each individual mathematical domain", and in a provocative tone, Arnold when referring to the International Conference of Mathematicians in Zürich in 1994,

> "In the first half of the century, the devil was winning. A special 'axiomatic-bourbakist' method of exposition of mathematics was invented, to make it perfectly pure (...) This formal proof is urged on students by the criminal bourbakizers and algebraizers of mathematics."[38]

The wish is clearly to break with the monolithic approach of knowledge founded solely on itself and projecting an image of isolation on the part of mathematical doing with regard to other disciplines. This desire affects both those disciplines which take on a greater

[36] LANG, S., *Serge Lang fait des maths en public. 3 Débats Au Palais de la Découverte*, Belin, Paris, 1984. Translator into Spanish P. J. Salas: *El placer estético de las Matemáticas*, Alianza Editorial, Madrid, 1992, p. 9.

[37] LANG, S., *El placer estético de las Matemáticas*, p. 9.

[38] ARNOLD, V., "Will Mathematics Survive?", *Mathematical Intelligencer*, v. 17, n. 3, (1995), p. 8.

role as well as the criteria of rigour itself and above all the exclusively deductivist and enunciative nature of mathematical doing. Analogy comes to the fore once more and we accept questions which are only hinted at by these analogies, conjectures which bring together apparently unrelated disciplines such as algebraic geometry, theory of singularity, combined topology of convex polynomials...

Rigour becomes a social criteria and the axiomatic-deductive nature was shown to be, in practice, mere fiction as no mathematical discipline fully reaches that level but only approaches it to a greater or lesser degree, without that gradualism having been clearly defined at any moment. Moreover, the role of ideas and physical concepts poses new problems as to how far we can accept "intuitive" results from Physics, which are not proved from the point of view mathematical principles. The theory of strings, nodes with Witten's contributions, Jones' invariants, the so-called R. Feynman integral, highly successful in many problems despite its not being sustained by any rigorous definition..., are questions to be discussed in this area.

Special mention must be made of course of the computer which, as Steen has pointed out, has changed the face of the mathematical ecosystem, not so much following the proof of the four colours conjecture, in the possible change in what is understood by concept of proof. The computer has highlighted the role of conjecture as well as the possibility of visualising concepts or processes based on iteration. It has focused attention on the role of what has come to be known as mathematical "experiment". Building algorithms —whether random or otherwise—, the development of computational mathematics with its numerical calculus, its application to solving equations in partial derivatives..., with its problems related to the complexity of the solution of mathematical problems, and above all the conjecture of the complexity or otherwise of polynomials, have aroused interest. In 1991 a journal was created with the title *Experimental Mathematics* in which, by way of a declaration of intention among other points, it states that mathematicians

"are not only interested in theorems and proofs but also in the way they have been or might be obtained."[39]

3.4.1. Peter Lax

If in previous cases I have cited certain mathematicians by way of an example, here the choice is Peter Lax, recipient of the Third Abel Prize in Mathematics, in 2005, who follows the new image of the mathematician essentially reflected by his fellow countryman J. von Neumann. In 1986, he writes

> "Applied and pure mathematics are more closely bound together today than any other time in the last 70 years."[40]

Lax recognises the seminal role of the computer in mathematical doing. In 1989 he stated

> "High-speed computers' impact on mathematics, both applied and pure, is comparable to the role of telescopes in astronomy and microscopes in biology."[41]

---

[39] LAX, P., "Mathematics and its Application," *Mathematical Intelligencer*, v. 8, n. 4, (1986), p. 14.

[40] LAX, P., "Mathematics and its Application," p. 14

[41] LAX, P., "The Flowering of Applied Mathematics in America," in DUREN, P., ASKEY, R. A., and MERZBACH, U. C. (eds.), *A Century of Mathematics in America*, American Mathematical Society, part II, Providence, RI, 1989, p. 465.

In contrast to bourbakist aesthete purism —reflected by Lang's cited words—, Lax has attempted to maintain his work in the interaction between mathematics and physics, now through the use of the computer. In the field of classical analysis, his work focuses on the theory of differential equations, a basic topic for mathematical applications, where each kind of differential equation requires, in general, its own methods for possible solutions. For Lax it is not enough to merely provide the conditions for an equation to have a solution in addition to proving that this is unique; one must put forward a value for this solution, which is not always feasible unless one uses the computer. In fields such as hyperbolic differential equations, present in shockwaves for example, Lax establishes what is known as the *Lax-Milgram theorem* for the existence of a solution and, for uniqueness, *Lax's Equivalence Theorem*; while to compute the solution he introduces numerical schemes in the case of laws of hyperbolic conservation such as the so-called *Lax-Friedrichs* and *Lax-Wendroff Schemes.*

Where his work has perhaps had the greatest repercussion has been in the "single wave" problem, with the model of the Korteberg-de Vries equation, the KdV equation. The use of numerical simulations led to the discovery that KdV solutions interacted like particles —hence the name solitons— and Lax, focusing on the properties of invariance of associated lineal problems, established through so-called *Lax pairs*, the internal mechanism of the inverse transform of the Lax-Phillips scattering theory. This field of work has had enormous consequences ranging from quantum and relativist theories to automorphic functions or communication by fibre optics...

The link between pure and applied mathematics to which we may now add the use of the computer, Lax reminds us,[42] leads to achievements such as the 1960s physicist Cormack being awarded the Nobel Prize for the creation of Computerised Tomography, the key to which lay in the development of a practical method to determine a function from its integrals along straight lines. Peter Lax insists on young talented mathematicians being included in this line of work. This mathematical doing reveals its power for the study of physis through simulation, creation of models, computation..., in addition to classical development. Mathematics does not appear as mere formal derivative syntax, but shows itself to be instrumental without surprises due to its unreasonable effectiveness...

3.4.2. Reflexive Judgement

At the current time we lack sufficient perspective to offer a calm analysis. The monolithic axiomatic-structuralist approach has been succeeded by something of a divergence of standpoints with their accompanying criticisms and controversies, such as those caused by the expressions of V. Arnold, or by A. Jaffe and F. Quinn with regard to the physical intuitive proofs, or those linked to the role of the computer. A time of crisis which, as always, mirrors the permanent state of mathematical doing, as explicitly recognised by the words of Lebesgue cited above and which the examples mentioned only serve to underline. A period of crisis in which mathematical doing strives to regain one of its own domains: that of being something which is truly done, in other words an art or technique with its normative role and associated value and not merely the reflection of a theoretical essentialist element or accumulation of axiomatised formal theories.

[42] Cf. LAX, P., "Mathematics and Computing," in ARNOLD, V., ATIYAH, M., LAX, P., and MAZUR, B. (eds.), *Mathematics: Frontiers and Perspectives*, American Mathematical Society, Providence, RI, 2000, pp. 417-432.

Abandoning the only approach of a fundamentalist nature, and returning to the quest for constructive norms or rules enables us to rid ourselves of the clash between the ideal and the real. Avoiding the dominance of discourse over doing opens up the door to a knowledge of physis. This involves adopting an approach where the propositions or principles we wish to reach —not those from which we start— are considered the governing or normative and not veritative, principles, which characterise the field of play yet without marking the boundaries in any clear way, as these principles may vary during the course of our work.

This, as a theoretical product and vis-à-vis the properties of concepts although not in terms of problems and solutions, does not prevent us from searching for rigour which may be provided to us not by the computational or what is strictly provable, but through what referring to Kant, may be termed as *reflexive judgement.* It is the "philosophical" expression of the process by which Laplace, as I cited, states that we are able to reveal the metaphysics of the methods that serve mathematicians. A reflexive judgement in which we start from the given particular —the problem to be solved, the proposition to be proved, the concept vaguely hinted at, the glimpse of an analogy...— to reach the general from whence we return to now state that the general we have reached is the determining principle which subsumes the given particular, principle or starting point of constructive mathematical doing.

Reflexive judgement allows us to cast to one side the split between pure and applied mathematics and pave the way to positivities where the reflexive opens up to the realm of what is possible. This realm with its attempts at conceptual precision and corresponding search for properties which may be attributed to such notions, properties that make up the theorems which, half emerging, must be proved, a proof which, naturally, stems from what is glimpsed, from what we wish to prove and appears as a process of rigour to ensure what is hinted at, what is sensed but what can seldom be realised absolutely.

Today we face a situation from which classical ontological positions in the shape of Realism, Conceptualism, and Nominalism may be seen as dogmatic systems based on more of a search for an object than on the possibility of experience in reflexive judgement as the key to endow what is glimpsed with ontological support.

4. Summing Up

The examples I have outlined clearly indicate that the mathematician who is a "creator" is guided at every moment by an ideology, by what he considers *his* philosophy of mathematical doing. The mathematician does not ply his trade "blindly," nor is he extravagant when searching for the "metaphysics" in his work. He does so faced with other conceptions, ever striving to clarify and specify the concepts used, both at a theoretical level and when solving problems.

Consequently we may not talk of one definitive, monolithic and eternal Philosophy of Mathematics, but rather of *Philosophies* of Mathematics at a given moment and time and, above all, for a given mathematician. What he aims to do in his reflexive judgement is mathematical doing which, concurrently, reflects a "metaphysics," a metaphysics which is far more modest than any attempt to reach the final solution to all conceptual problems —with the exception of Hilbert and Brouwer, characteristic of the ideological and political fundamentalisms that ran through the twentieth century. At each point in his praxis, he is striving for a critical analysis of the conceptual elements he finds and which are not wholly and radically clear, like the demonstrative processes in the field in which he is immersed at each instant.

The philosopher might not be pleased with the admission that the only feasible philosophy is local and may wish to construct more comprehensive and abstract ideologies. When dealing with mathematics he will, despite his attempt at a generalist approach, be dependent on a specific context and mathematical fashion. This can be seen in the latest philosophies of mathematics: renewed logicisms, structuralisms, nominalisms *a la* Field, constructivism reliant on the latest results of non-canonical analysis, supposed empiricisms... Philosophies of Mathematics which rely on prior ideologies, and of course, on specific mathematical doing that lacked any sense in 18th century mathematical doing, for instance.

The philosophy of mathematics depends on what it attempts to philosophise about. To some extent, it must play servant to mathematical doing, which it wishes to analyse critically. If this doing changes and is transformed, critical analyses must also vary and take account of these changes. The door cannot be closed to new approaches, new ways of analysis, new content. Nor must this time-space analysis be unchanging, "a single thought". Not even the predominant image of 1950s Bourbakist formalism was able to erase mathematical doing with a more constructive approach aimed at maintaining its links to other disciplines, where the goals of each discipline were established at the end and not the beginning of the work, a line of work which has taken on an increasingly prominent role at the current time.

5. Bibliography

Abel, N. H., *Correspondance. Mémorial*, Kristiania, 1902.

Arnold, V., "Will Mathematics Survive?," *Mathematical Intelligencer*, v. 17, n. 3, (1995), pp. 6-10.

Arnold, V., Atiyah, M., Lax, P., and Mazur, B. (eds.), *Mathematics: Frontiers and Perspectives*, American Mathematical Society, Providence, RI, 2000.

Aroca, J. M., "Zariski y la Escuela italiana," en: *Historia de la Matemática en el s. XX*, Real Academia de Ciencias Exactas, Físicas y Naturales, Madrid, 1998, pp. 171-193.

Bernkofp, M., "Halphen, Georges Henri," *Dictionary of Scientific Biography*, vol. VI, Scribner's and Sons, New York, 1972, pp. 75-76.

Bertini, E., *Complementos de Geometría Proyectiva*, Aguilar, Madrid, 1951.

Borel, E., *Leçons sur la Théorie des Fonctions*, Gauthiers-Villars. Paris, 4th ed., 1950.

Botazzini, U. and Dahan-Dalmedico, A. D. (eds.), *Changing Images in Mathematics*, Routledge, London, 2001.

Brigaglia, A., "The Creation and Persistence of Natural Schools: The Case of Italian Algebraic Geometry," in Botazzini, and Dahan-Dalmedico, (eds), *Changing Images in Mathematics*, Routledge, London, 2001, pp. 187-206.

CAUCHY, A, *Curso de Análisis*, Colección Mathema, UNAM, Mexico D. F., 1994 (Selection and translation by Carlos Álvarez).

CARTIER, P., "The Continuing Silence of Bourbaki," *Mathematical Intelligencer*, v. 20, n. 1, (1998), pp. 22-28 (An Interview by Marjorie Senechal).

DAHAN-DALMEDICO, D., "An Image Conflict in Mathematics after 1945," in BOTAZZINI, U. and DAHAN-DALMEDICO, D. (eds.), *Changing Images in Mathematics*, Routledge, London, 2001, pp. 223-254.

DUGAC, P., *Histoire de l'Analyse*, Vuibert, Paris, 2003.

EWALD, W., *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, Vol. 2, Clarendon Press, Oxford, 1996.

FELIZ, L., *Message d'un mathématicien: Henri Lebesgue*, Albert Blanchard, Paris, 1974.

JAFFE, A. and QUINN, F., "Theoretical Mathematics: Toward a Cultural Synthesis of Mathematics and Theoretical Physics," *Bulletin of American Mathematical Society*, v. 29, (1993-1994), pp. 1-13. (Replies in v. 30, pp. 178-207.)

LAGRANGE, J. L., *Théorie des fonctions analytiques, contenant les principes du calcul différentiel, dégagés, de toute considération d'infiniment petits ou d'evanouissans, de limites ou de fluxions, et réduits a l'analyse algébrique des quantités finies*, l'Imprimerie de la République, Paris, 1797.

LANG, S., *Serge Lang fait des maths en public. 3 Débats Au Palais de la Découverte*, Belin, Paris, 1984. Translator into Spanish P. J. Salas: *El placer estético de las Matemáticas*, Alianza Editorial, Madrid, 1992. (English translation of the French text: *The Beauty of doing Mathematics. Three Public Dialogues*, Springer-Verlag, New York, 1985.)

LAX, P., "Mathematics and its Application," *Mathematical Intelligencer*, v. 8, n. 4, (1986), pp. 14-17.

LAX, P., "The Flowering of Applied Mathematics in America," in DUREN, P., ASKEY, R. A., and MERZBACH, U. C. (eds.), *A Century of Mathematics in America*, American Mathematical Society, Part II, Providence, RI, 1989, pp. 455-466.

LAX, P., "Mathematics and Computing," in ARNOLD, V., ATIYAH, M., LAX, P., and MAZUR, B. (eds.), *Mathematics: Frontiers and Perspectives*, American Mathematical Society, 2000, pp. 417-432.

LEBESGUE, H., *Les lendemains de l'intégrale. Lettres à Émile Borel*, Vuibert, Paris, 2004. (Preface by G. Choquet.)

LEIBNIZ, G. W., *Discours de métaphysique*, 1686. Introduction and translation by J. Marías: *Discurso de Metafísica*, Revista de Occidente, Madrid, 1942.

LORENZO, J. DE, *Filosofías de la Matemática fin de siglo XX*, Ediciones de la Universidad de Valladolid, Valladolid, 2000.

LUZIN, N., *Leçons sur les ensembles analytiques et leurs applications. Avec une note de W. Sierpinski*, Gauthier-Villars et Cie., Paris, 1930, Préface de Henri Lebesgue. Reprinted in Chelsea Publishing Company, New York, 1972. Preface by H. Lebesgue, pp. VII-XI.

MEDVEDEV, F. A., *Scenes from the History of Real Functions*, Birkhauser, Boston, 1991. Original Russian edition, Ed. Nauka, Moscow, 1975.

MERAY, CH., *Lecons nouvelles sur l'analyse infinitesimales et ses applications géométriques,* Gauthiers Villars, Paris, 1894.

POINCARE, H., "Sur les groupes des equations lineaires," *Acta Mathematica*, v. 4, (1884), pp. 201-312. Translated into English by J. Stillwell: POINCARÉ, H., *Papers on Fuchsian Functions*, Springer, New York, 1985.

ZARISKI, O. and SAMUEL, P., *Conmutative Algebra. Vol. I*, van Nostrand, N. York, 1958.

ZERMELO, E., "Beweis, daß jede Menge wohlgeordnet werden kann," *Mathematische Annalen*, v. 59, (1904), pp. 514-516.

Ontological and Epistemological Problems of Mathematics

Jesus Alcolea-Banegas[*]

"The mathematician is enthralled by the timeless self-identity of essences, and hence always gravitates spontaneously to one form of Platonism or another"

W. Barrett[**]

1. Introduction. Ontology and Epistemology in the Beginnings of Modern Philosophy of Mathematics

In this introduction we are going to describe some views espoused in contemporary discussions in the philosophy of mathematics, so as to get some sense of the subject.

The prescriptive view of mathematics tended to reduce it to the logic of the sciences, and such a tautological view overlooked the frequent use of complex numbers and other mathematical objects. Given their concern with consistency and epistemological foundations, for instance, the intuitionists rejected the use of infinity in mathematics, though it seemed to be necessary for the calculus. Such difficulties led people like Carnap to take the pragmatic view that we use mathematical systems because they are effective, and that we should not endow them with any real ontological status. If the pragmatic view were correct, however, it would mean that we can adopt as many different mathematical systems as suit our purposes, whereas the tendency is towards greater unification and simplification.

Modern intuitionists made the existence of mathematical objects depended on some construction or finite procedure like counting. But the basic standpoint of modern intuitionists was that classical mathematics employed forms of reasoning which are not valid on any legitimate way of construing statements. Thus they tended to assume that the meaning of a mathematical statement is exhaustively determined by use rather than by reference to some independently existing set of objects. In contrast, a platonistic interpretation of a mathematical statement took the central notion to be that of truth as correspondence with such objects. Given the theory of meaning which underlies platonism,[1] our grasp of the meaning of a mathematical statement consists in our knowledge of the conditions required for it to be true, although we cannot in general recognize when such conditions obtain. For instance, when the statement is effectively undecidable, as when the infinite is involved, the conditions which must obtain for it to be true are not accessible to us. So the meaning of such a statement can only be grasped by implicit knowledge that is not exhausted by the capacity to state what is known.

[*] The author is grateful to the Spanish Ministry of Science and Education for supporting partially this work (Project: HUM2005-00365/FISO). He is also grateful to W. J. González, J. de Lorenzo, and D. Gillies for their commments on this paper.

[**] Barrett, W., *Irracional Man. A Study in Existential Philosophy*, Doubleday Anchor Books, Garden City, NY, 1962, p. 106.

[1] In what follows, although we will have the chance to speak a lot about 'platonism', it seems to be in order to point out that the term was introduced by Paul Bernays referring to a generalized tendency that, in his opinion, was common in the mathematics of his time. Cf. Bernays, P., "On Platonism in Mathematics," (1935), in Benacerraf, P. and Putnam, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, Cambridge University Press, Cambridge, 1983, pp. 258-271.

However, on the assumption that use exhaustively determines meaning, the ascription of implicit knowledge is meaningful only when someone can fully manifest that knowledge in suitable circumstances. Thus the platonistic theory of meaning cannot be one in which meaning is fully determined by use. But if to know the meaning of a mathematical statement is to grasp its use, then proof rather than truth becomes central for a theory of meaning of mathematical statements. In other words, grasping the meaning of a statement consists in a capacity to recognize a proof of it when it is presented. In the case of statements about the infinite, however, such a proof will always consist in finite procedures which undermine the truth of any claims about an actual infinite. So modern intuitionists talked about a potential infinite.

We can see this clearly if we consider the intuitionistic thesis that mathematical statements do not relate to an objective mathematical reality existing independently of us. In other words, mathematical objects are creations of the human mind whose mode of being depends on being thought. Let us note that such a thesis is quite compatible with a realist view concerning the physical universe and statements about its physical properties. A conception of meaning as determined by truth-conditions is available for any statements which relate to an independently existing reality, since we can assume that each statement has a determinate truth-value. But when the statements do not relate to such an external reality, the supposition that each of them has a determinate truth-value is empty, and so their meaning cannot be given by their truth-conditions.

According to Dummett,[2] the difficulty in modern philosophy of mathematics concerns how we are to decide whether mathematical objects are the creations of human thought before deciding what is the correct model for the meanings of mathematical statements. One is faced with the ontological question of whether mathematical objects are human creations (constructivism) or whether they are independently existing abstract objects (platonism). Crudely stated, the platonist picture likens mathematical inquiry to astronomy, which investigates mathematical structures just like galaxies that exist in another realm which we do not inhabit but which we can observe and report on.[3] By contrast, the constructivist picture likens mathematical activity to that of a craftsman fashioning objects through his imagination.

But, if Frege is right, we cannot refer to any object except by saying something about it. Hence any thesis about the ontological status of certain kinds of objects must also involve a thesis about the truth of statements referring to such objects. In other words, we cannot separate the question of the ontological status of a class of objects from the question of the correct notion of truth for statements about those objects. As we have suggested, the correspondence theory of truth tends to support platonism, while intuitionism relies on the notion of proof and its associated truth-conditions. But substantial disputes between these two models of meaning only emerge over statements that are effectively undecidable; for

[2] Cf. Dummett, M., "The Philosophical Basis of Intuitionistic Logic," (1975), in Benacerraf, P. and Putnam, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, p. 110.

[3] Quite dogmatically, the famous mathematician G. H. Hardy believed that mathematical reality has an external existence independent of mathematicians who have as an aim "to discover it or *to observe it* and that theorems … are only the products of our observations," Hardy, G. H., *A Mathematician's Apology*, Cambridge University Press, London, 1940. Translated into Spanish: Hardy, G. H., *Autojustificación de un matemático*, Ariel, Barcelona, 1981, p. 121.

instance, those involving quantification over infinite totalities like the natural numbers. Whereas platonists are willing to accept such statements as true or false, intuitionists deny that the logical law of the excluded middle applies to incompletable procedures.

This tension between truth conditions and knowledge in contemporary philosophy of mathematics signals a more fundamental problem about the existence of mathematical objects like transfinite sets. According to platonists like Frege, who tried to reduce arithmetic to logic, the existence of an actual totality of integers is a necessary assumption for the logical principle of the excluded middle to be applicable to number theory. Similarly, the impredicative definition of a real number depends on assuming that the totality of integers exists, since a real number is a Dedekind cut involving an infinite series of rational numbers. But such assumptions have been undermined by the notorious paradoxes of set theory and by the vicious circularity associated with impredicative definition.

In order to avoid the difficulties of extreme platonism, Carnap applied constructivity consisting in the explicit definition of mathematical notions from the basic logical concepts,[4] as opposed to their introduction by means of axioms, and in the rule that in mathematics only what has been proved in finitely many steps may be taken to exist. This brought him some way towards the intuitionist position that every logical-mathematical operation or proof or definition must be finite. But intuitionists like Heyting went much further by claiming that the existence of mathematical objects and their properties is guaranteed only insofar as they are determined by human thought.[5] Against such anthropological mathematics,[6] however, the formalists objected that the human mind does not have exact images of straight lines or of very large numbers,[7] so that these objects cannot depend on thought for their existence. Of course, they conceded that the mathematician can methodically develop a series of relations for these objects, yet they insisted that these relations are independent of the mental significance we attach to them. Such a meaningless series of relations gains mathematical existence only when it has been represented in spoken or written language, together with the mathematical laws upon which its development depends. Hilbert insisted that the infinite as a completed totality is an illusion that must be replaced by finitist procedures that give the same results in mathematical deductions. He pointed out that the infinite divisibility of the continuum, for instance, is an operation that exists only in thought, since modern physics shows that nature does not provide a suitably homogeneous continuum.

In Brouwer's case intuitionism represented an ingenious solution to a problem inherited from Kant, who minimized the role of discursive arguments in mathematics because he thought that the basic axioms and every step of a proof were grounded in pure intuition. Brouwer's solution was to distinguish between mathematical activity as mental construction

---

4 CARNAP, R., "The Logicist Foundations of Mathematics," (1931), in BENACERRAF, P. and PUTNAM, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, pp. 41-51.

5 Cf. HEYTING, A., "The Intuitionist Foundations of Mathematics," (1931), in BENACERRAF, P. and PUTNAM, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, pp. 52-60.

6 As was named by CARNAP, R., "The Logicist Foundations of Mathematics," p. 49.

7 Cf. HILBERT, D., "On the Infinite," (1925), in BENACERRAF, P. and PUTNAM, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, pp. 183-201; and VON NEUMANN, J., "The Formalist Foundations of Mathematics," (1931), in BENACERRAF, P. and PUTNAM, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, pp. 61-65.

based on the pure intuition of time, and its linguistic expression or communication. Thus he resolved the epistemological problem about the source of mathematical certainty by appealing to intuition, while explaining the nature of mathematical proof as a construction of constructions, each of which is ultimately based on intuition. Similarly, ontological problems about the nature of mathematical objects and their mode of existence were resolved by Brouwer with a doctrine which had two sides: constructivism and mentalism.[8] In other words, he held mathematical objects to be nothing more than constructions of the human mind, whose objectivity depends entirely on repeating their construction. By contrast with formalists like Hilbert, for whom exact mathematical objects exist only on paper, the intuitionists located such objects in the human intellect.

A brief review of conflicting views on the question of mathematical existence shows how difficult it is even to find common ground for an answer. Thus empiricists tended to reject abstract objects, while giving a nominalist account of language that purportedly refers to them. For instance, in the case of a mathematician who talks about infinite numbers and sets, they would insist that s/he is talking about meaningless symbols which are manipulated according to formal rules.[9] But the old problem about abstract objects has been given new currency by fresh developments in semantic theories of meaning and truth. While some semanticists hold that scientific expressions designate not only concrete material things but also abstract objects, others question that such claims go beyond empirical science into the realm of platonist ontology. From a philosophical perspective, however, it is difficult to mediate between ontological views that differ so radically as the empirical assumption that only sensible physical things exist, and the platonist view that there are abstract objects.

One modern way of eliminating the problem is through Carnap's distinction between internal and external questions about existence.[10] The former are questions about the existence of new kinds of objects within a given framework, which may be routinely answered either by logical or empirical methods, depending on the character of the framework. For example, the question of whether there is a prime number greater than 100 can be answered through logical analysis based on the rules of the number system. In contrast, the question of whether or not numbers exist was problematic for Carnap because it is unclear whether it is an internal or an external question. It can hardly be an internal question, however, since that would amount to nothing more than whether or not the framework of numbers is empty of content. So he inferred that it must be taken as an external question about the existence of the system of objects as a whole; i.e. the question of whether or not numbers have an ontological character called "reality" or "subsistence." Whichever way one formulates the question, however, Carnap found it to be meaningless because it is not formulated in the common scientific language and so lacks cognitive content. Therefore, he thought it as a pseudo-question.

---

[8] In fact, the evolution of Brouwer's thought led him to a sort of solipsism. Cf. Brouwer, L. E. J., "Consciousness, Philosophy, and Mathematics," (1949), Benacerraf, P. and Putnam, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, pp. 90-96; and Alcolea, J., "Un aspecto del intuicionismo matemático," *Quaderns de Filosofia i Ciència*, n. 9-10, (1986), pp. 13-20.

[9] David Hilbert could be assigned to this position.

[10] Cf. Carnap, R., "Empiricism, Semantics and Ontology," (1950), in Benacerraf, P. and Putnam, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, pp. 241-257.

Carnap's argument could be considered as a typical example of the radical challenge to traditional ontology that was posed by the logical positivists. Despite their dislike for ontology, however, they had to give some explanation of the reference to abstract objects to be found in the language of mathematics and physics, which they placed on the same methodological level. Carnap said also that the acceptance of a new linguistic form cannot be judged as being either true or false,[11] since it is not an assertion, but only as being more or less expedient or fruitful. Against Quine's notion of ontological commitment,[12] Carnap insisted that the acceptance of a linguistic framework must not be taken to imply any ontological doctrine concerning the reality of the objects in question. But Quine had his own answer:

> "Carnap's separation of questions of existence into questions of fact and questions of framework was a separation of the synthetic and the analytic. Collapse this epistemological duality and you collapse the ontological duality. Sticks, stones, sets, and numbers all become, for me, denizens of the world on equal footing. Values of variable."[13]

And that collapse was produced in his paper "Two Dogmas of Empiricism."[14]

As another opponent to positivists, Popper concedes that foundational questions about the sciences involve an ontological dimension when he posits a so-called "World 3,"[15] consisting of objective contents of knowledge. Although he denies that this is the same as Plato's theory of Forms,[16] he does seem committed to *some kind of* platonism when he likens his World 3 to Frege's objective content of thought.[17] In some ways, Popper's approach to the physical sciences is similar to that of Frege to the mathematical sciences. For instance, just as Frege criticized accounts which appealed to psychological states, so also Popper resists the subjective psychologism in epistemology which was

---

[11] Cf. Carnap, R., "Empiricism, Semantics and Ontology," p. 250.

[12] Cf. Quine, W. v. O., "On What There Is," (1948), in Quine, W. v. O., *From a Logical Point of View*, Harper & Row, New York, 2nd. edition revised, 1961, pp. 1-19.

[13] Quine, W. v. O., "Two Dogmas in Retrospect," *Canadian Journal of Philosophy*, v. 21, n. 3, (1991), p. 271.

[14] Quine, W. v. O., "Two Dogmas of Empiricism," (1951), in Quine, W. v. O., *From a Logical Point of View*, pp. 20-46.

[15] Cf. Popper, K. R., "Epistemology Without a Knowing Subject," (1968), in Popper, K. R., *Objective Knowledge*, Clarendon Press, Oxford, 1972, pp. 106-152; and Popper, K. R., "On the Theory of the Objective Mind," (1972), in Popper, K. R., *Objective Knowledge*, pp. 153-190.

[16] Interestingly enough, Popper wrote: "Plato's third world was divine; it was unchanging and, of course, true. Thus there is a big gap between his and my third world: my third world is man-made and changing," Popper, K. R., "Epistemology Without a Knowing Subject," p. 122. And "… it seems that [Plato] did not realize that the third world contained not only universal concepts or notions, such as the number 7 or the number 77, but also mathematical truths or propositions, such as the proposition '7 times 11 equals 77', and even false propositions, such as '7 times 11 equals 66', and, in addition, all kinds of non-mathematical propositions or theories," Popper, K. R., "On the Theory of the Objective Mind," p. 156.

[17] Let us remember that Frege marked a difference between "objective" and "actual": "I distinguish what I call objective from what is handleable or spatial or actual. The axis of the earth is objective, so is the centre of mass of the solar system, but I should not call them actual in the way the earth itself is so," Frege, G., *The Foundations of Arithmetic* (1884), Blackwell, Oxford, 2nd. revised edition, 1953, p. 35.

In another place, Frege seems to come very close to introducing abstract objects by analogy with physical objects, because both thoughts and numbers are non-physical, objective objects: "... thoughts are neither things in the external world nor ideas. A third realm must be recognized. Anything belonging to this realm has it in common with ideas that it cannot be perceived by the senses, but has it in common with things that it does not need an owner so as to belong to the contents of his consciousness," Frege, G., *Collected Papers on Mathematics, Logic, and Philosophy*, edited by B. McGuinness et al., Blackwell, Oxford, 1984, p. 363.

typical of post-Cartesian "belief-philosophers". He does this by making a distinction between knowledge or thought in the subjective sense (World 2) and in an objective sense (World 3). While he regards the first sense of knowledge as being largely irrelevant for a theory of knowledge, Popper claims that every epistemology must concern itself with the second sense of knowledge "without a knowing subject." In this respect, therefore, Popper is justified in calling his position "realism". In fact, "though originally constructed by us," mathematical objects may become a part of World 3.[18] But surely mathematics is objective and at the same time abstract: a whole world of problems and solutions that we do not invent, but we rather discover.

What should be clear from this survey of some issues in contemporary philosophy of mathematics is that the post-Cartesian assumptions about the primacy of the knowing subject which govern modern philosophy dictate that epistemology takes precedence over ontology. But the debate continues.

2. Three Important Arguments and the Debate about Platonism and Anti-Platonism in Philosophy of Mathematics

For some time now ontological debates in the philosophy of mathematics have been dominated by three relevant arguments. While the first of these arguments is usually taken to be an argument for platonism, the other two are taken to raise problems for platonism:

(1) The Quinean argument that mathematical objects are indispensable to our best physical theories and therefore share the ontological status of scientific objects.[19]

(2) The Benacerraf indeterminacy objection to the natural numbers being identified with sets.[20] For instance, should we identify the natural number 2 with von Neumann's set {Ø, {Ø}} or with Zermelo's set {{Ø}}?

(3) The dilemma posed by Benacerraf[21]: if mathematical objects such as functions, numbers, and sets have mind-independent, though admittedly abstract, existence, how is it that we have knowledge of them?

Before discussing the arguments in more detail, some preliminaries are in order, since they are supposed to have serious implications for the debate about platonism and anti-platonism in mathematics. There are many different ways to characterize them. Perhaps the most common way is —as we have just suggested— as a thesis about the existence or non-existence of mathematical objects. Thus, according to this conception of platonism, mathematical objects such as functions, numbers, and sets have mind- and language-independent existence or, as it is also commonly expressed, we *discover* rather than invent mathematical theories (which are taken to be a body of facts about the relevant mathematical objects). This is usually called *ontological platonism*. Anti-platonism, then,

[18] "Epistemology Without a Knowing Subject," p. 138.

[19] Cf. Quine, W. v. O., "Two Dogmas of Empiricism," pp. 44-46.

[20] Cf. Benacerraf, P., "What Numbers Could Not Be," (1965), in Benacerraf, P. and Putnam, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, pp. 272-294.

[21] Benacerraf, P., "Mathematical Truth," (1973), in Benacerraf, P. and Putnam, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, Cambridge University Press, Cambridge, 1983, pp. 403-420.

is the position that mathematical objects do not enjoy mind-independent existence or, alternatively, we *invent* rather than discover mathematical theories.

Another common characterization of platonism or realism is via truth. For instance, Dummett says that realism is "the belief that statements of the disputed class possess an objective truth-value, independently of our means of knowing it: they are true or false in virtue of a reality existing independently of us."[22] This is usually called *semantic realism*. Putnam, another philosopher who prefers this way of characterizing realism, points out that, according to this view, it is possible to be a mathematical realist without being committed to mathematical objects: realism is about objectivity, and not necessarily about objects[23], following Kreisel.[24]

Certainly one reason many philosophers and mathematicians find appealing the view that mathematical objects exist is that it does provide mathematics with the desired objectivity. For instance, Gödel was drawn to this position for that very reason. Platonism purported to provide the independent questions of mathematics with objective answers:

> "The truth, I believe, is that these concepts form an objective reality of their own, which we cannot create or change, but only perceive and describe." So, "... the Platonistic view is the only one tenable. Thereby I mean the view that mathematics describes a non-sensual reality, which exists independently both of the acts and [of] the dispositions of the human mind and is only perceived, and probably perceived very incompletely, by the human mind."[25] In fact, "set-theoretical concepts and theorems describe some well-determined reality," although "the axiomatic system of set theory as used today is incomplete, [and] it can be supplemented without arbitrariness by new axioms which only unfold the content of the concept of set."[26]

If this were all there was at issue, then Putnam's account of realism would suffice, but objectivity is not the only issue. The important question in the ontology of mathematics is: Do mathematical objects exist? Many philosophers have taken this question to be interesting and worth pursuing, particularly since the seminal papers by Quine and Benacerraf.

We should point out that the terms "mathematical realism" and "platonism" (or "Platonism") have been used interchangeably in modern literature. With platonism has been identified the view that (a) mathematical objects exist and, what is more, that their existence is not mind or language dependent, and (b) mathematical statements are true or false in virtue of the properties of these mathematical objects. But sometimes "platonism" has been used to imply that mathematical objects are causally inert, that they are not located in space-time, or that they exist necessarily. Although such views are typically

---

[22] DUMMETT, M., "Realism (1963)," in DUMMETT, M., *Truth and Other Enigmas*, Duckworth, London, 1978, p. 146.

[23] PUTNAM, H., "What is Mathematical Truth?," in PUTNAM, H., *Mathematics, Matter and Method*, Cambridge University Press, Cambridge, MA, 2nd. edition, 1979, pp. 69-70.

[24] KREISEL, G., "Wittgenstein's Remarks on the Foundations of Mathematics," *The British Journal for the Philosophy of Science*, 9 (1958-59), pp. 135-158.

[25] GÖDEL, K., "Some Basic Theorems on the Foundations of Mathematics and their Philosophical Implications (1951)," in GÖDEL, K., *Collected Works*. Vol. III. *Unpublished Essays and Lectures*, edited by S. Feferman et al., Oxford University Press, Oxford, 1995, pp. 320 and 322-323.

[26] GÖDEL, K., "What is Cantor's Continuum Problem?," (1964), in GÖDEL, K., *Collected Works*. Vol. II. *Publications 1938-1974*, edited by S. Feferman et al., Oxford University Press, Oxford, 1990, pp. 260-261.

endorsed by platonists, such endorsement is by no means universal. For instance, Bigelow and Maddy have both put forward accounts of mathematics,[27] which they describe as platonist accounts, in which mathematical objects are located in space-time and are part of the causal nexus. Others propose accounts where mathematical objects exist contingently, that even all objects are abstract objects, that we can have empirical knowledge of various abstract objects, and that we might causally interact with them.[28] All these accounts postulate mathematical objects and, from an ontological point of view, this is what matters. We will thus use the term "platonism" to include all of them.

In the opposite side, any account that denies that mathematical objects exist is anti-realist. Here the nominalist positions must be included,[29] and Field could be an instructive advocate.[30] His basic argument is more or less like this: the function of mathematics within empirical science is to assist in formulating scientific laws and deducing scientific statements from each other. If this use of mathematics can be shown to be dispensable, then the truth of mathematics will not be presupposed by science, and thus Quine's indispensability argument for the reality of mathematical objects will be undercut. Field tries then to establish a reformulation of science (physics) which will avoid reference to those objects. Field has also developed a sort of fictionalism,[31] according to which mathematical statements are, strictly speaking, untrue; they are true only in the sense that they are true of the story of mathematics:

> "... the fictionalist can say that the sense in which "2 + 2 = 4" is true is pretty much the same as the sense in which "Oliver Twist lived in London" is true: the latter is true only in the sense that it is true according to a certain well-known story, and the former is true only in that it is true according to standard mathematics."[32]

There are various platonist and anti-platonist strategies in the philosophy of mathematics. Each of these has its own particular strengths and weaknesses. On one hand, platonist accounts of mathematics generally have the problems of providing an adequate epistemology for mathematics and of explaining the apparent indeterminacy of number terms. On the other, anti-platonist accounts generally have trouble providing an adequate treatment of the wide and varied applications of mathematics in the empirical sciences. There is also the challenge for anti-platonism to provide a uniform semantics for mathematics and other discourse.[33]

In recent times many platonist strategies have responded to the epistemological challenge by placing mathematical objects firmly in the physical realm. Thus Maddy

---

[27] Cf. Bigelow, J., *The Reality of Numbers. A Physicalist's Philosophy of Mathematics*, Clarendon Press, Oxford, 1988; and Maddy, P., *Realism in Mathematics*, Clarendon Press, Oxford, 1990.

[28] For a nice argument, cf. Tymoczko, T., "Mathematics, Science and Ontology," *Synthese*, v. 88, (1991), pp. 201-228.

[29] One might consult Burgess, J. and Rosen, G., *A Subject with No Object, Strategies for Nominalistic Interpretation of Mathematics*, Clarendon Press, Oxford, 1997, for an extended discussion and assessment of nominalist strategies in mathematics. After finishing this article, the author got Shapiro, S. (ed.), *The Oxford Handbook of Philosophy of Mathematics and Logic*, Oxford University Press, Oxford, 2005. Readers are invited to consult that great overview of the major problems and positions.

[30] Field, H., *Science Without Numbers. A Defence of Nominalism,* Blackwell, Oxford, 1980.

[31] Field, H., *Realism, Mathematics, and Modality*, Blackwell, Oxford, 1989, pp. 1-10.

[32] *Realism, Mathematics, and Modality*, p. 3.

[33] Cf. Benacerraf, P., "Mathematical Truth," p. 408.

argued that we can see sets.[34] When we see six eggs in a carton we are seeing *the set* of six eggs. This account provides mathematics with an epistemology consistent with other areas of knowledge by giving up one of the core doctrines of traditional platonism, that mathematical objects are abstract. In response to the apparent indeterminacy of the reduction of numbers to sets, one popular platonist strategy is to identify a given natural number with a certain position in *any* ω-sequence. Thus, it does not matter that 2 can be represented as {{Ø}} in Zermelo's ω-sequence and {Ø, {Ø}} in von Neumann's ω-sequence. What is important according to this account is that the structural properties are identical. This view is usually called *structuralism* since it is the structures that are important, not the items that constitute the structures.[35]

These are not meant to be anything more than cursory sketches of some of the available positions. But to finish let us quote some bold and interesting conclusions that Balaguer has obtained:

> "*Strong epistemic conclusion*: It's not just that we *currently* lack a cogent argument that settles the dispute over mathematical objects —it's that we could *never* have such an argument.
>
> *Metaphysical conclusion*: It's not just that *we* could never discover whether platonism or anti-platonism is true —it's that there is no fact of the matter as to which of them is true. In other words, there is *no fact of the matter* as to whether there exist any abstract objects."[36]

That is, since both platonism and anti-platonism are defensible and there are no good arguments against either, we do not have good reasons to think that one is better than the other. So, let us discuss briefly the three famous arguments.

(1) Quine's Indispensability Argument[37]

One of the most intriguing features of mathematics is its applicability to empirical science. Every branch of science draws upon large and often diverse portions of mathematics. Not only does mathematics help with empirical predictions, it allows elegant and economical statement of many theories. Indeed, so important is the language of mathematics that it is hard to imagine how some theories could even be stated without it. Furthermore, looking at the world through mathematical eyes has, on more than one

[34] Cf. *Realism in Mathematics*, pp. 58-63.

[35] Cf. Resnik, M. D., *Mathematics as a Science of Patterns*, Clarendon Press, Oxford, 1997; and Shapiro, S., *Philosophy of Mathematics. Structure and Ontology*, Oxford University Press, Oxford, 1997.

[36] Balaguer, M., *Platonism and Anti-Platonism in Mathematics*, Oxford University Press, Oxford, 1998, p. 17.

[37] Or *Quine-Putnam's indispensability argument*, because Putnam also once endorsed it: "... quantification over mathematical entities is indispensable for science, both formal and physical; therefore we should accept such quantification; but this commits us to accepting the existence of the mathematical entities in question. This type of argument stems, of course, from Quine, who has for years stressed both the indispensability of quantification over mathematical entities and the intellectual dishonesty of denying the existence of what one daily presupposes," Putnam, H., "Philosophy of Logic," (1971), in Putnam, H., *Mathematics, Matter and Method*, Cambridge University Press, Cambridge, MA, 2nd. edition, 1979, p. 347.

"In [1971] I argued in detail that mathematics and physics are integrated in such a way that it is not possible to be a realist with respect to physical theory and a nominalist with respect to mathematical theory," Putnam, H., "What is Mathematical Truth?," p. 74.

occasion, facilitated enormous breakthroughs in science. Indispensability arguments purport to yield conclusions about ontology based on this simple, undeniable fact. Indispensability arguments about mathematics urge platonists to place mathematical objects in the same ontological boat as other theoretical objects. That is, it invites them to embrace platonism. So Quine goes from "utility" to "irredemption":

> "A platonistic ontology of this sort is, from the point of view of a strictly physicalistic conceptual scheme, as much a myth as that physicalistic conceptual scheme itself is for phenomenalism. This higher myth is a good and useful one, in turn, in so far as it simplifies our account of physics. Since mathematics is an integral part of this higher myth, the utility of this myth for physical science is evident enough."[38]

> "Certain things we want to say in science may compel us to admit into the range of values of the variables of quantification not only physical objects but also classes and relations of them; also numbers, functions, and other objects of pure mathematics. For mathematics —not uninterpreted mathematics, but genuine set theory, logic, number theory, algebra of real and complex numbers, differential and integral calculus, and so on— is best looked upon as an integral part of science, on a par with the physics, economics, etc., in which mathematics is said to receive its applications. (...) We have reached the present stage in our characterization of the scientific framework not by reasoning a priori from the nature of science qua science, but rather by seizing upon traits of the science of our day."[39]

> "Ordinary interpreted scientific discourse is as irredeemably committed to abstract objects —to nations, species, numbers, functions, sets— as it is to apples and other bodies. All these things figure as values of the variables in our overall system of the world. The numbers and functions contribute just as genuinely to physical theory as do hypothetical particles."[40]

Quine draws attention to the fact that abstract objects, in particular mathematical objects, are as indispensable to our scientific theories as the theoretical objects of our best physical theories. But he suggests that anyone who is a platonist about theoretical objects but anti-platonist about mathematical objects is guilty of holding "a double standard" that he rejects.[41] Quine is also claiming that those portions of mathematical theories that are employed by empirical science enjoy whatever empirical support the scientific theory as a whole enjoys. Thus it seems reasonable to take science, or at least whatever the goals of science are, as the purpose for which mathematical objects are indispensable.

It will be useful here to outline the argument from naturalism and holism that gives support to indispensability. Naturalism, for Quine at least, is the philosophical doctrine that requires the "abandonment of the goal of a first philosophy" and that the philosophical

---

[38] Quine, W. v. O., "On What There Is," p. 18.

[39] Quine, W. v. O., "The Scope and Language of Science," (1957), in Quine, W. v. O., *The Ways of Paradox and Other Essays*, Cambridge University Press, Cambridge, MA, revised and enlarged edition, 1976, p. 244.

[40] Quine, W. v. O., *Theories and Things*, Harvard University Press, Cambridge, MA, 1982, pp. 149-150.

[41] Quine, W. v. O., "Two Dogmas of Empiricism," p. 45. As we saw previously, this is Carnap, who "has recognized that he is able to preserve a double standard for ontological questions and scientific hypotheses only by assuming an absolute distinction between the analytic and the synthetic; and I need not say again that this is a distinction which I reject," "Two Dogmas of Empiricism," pp. 45-46.

enterprise is continuous with the scientific enterprise. What is more, science, thus construed, that is, with philosophy as a continuous part, is taken to be the complete story of the world. This doctrine arises out of a deep respect for scientific methodology and an acknowledgment of the undeniable success of this methodology —observation and the hypothetico-deductive method— as a way of answering fundamental questions about all nature of things. As Quine suggests, one of its sources lies in the "unregenerate realism, the robust state of mind of the natural scientist who has never felt any qualms beyond the negotiable uncertainties internal to science."[42] For the metaphysician this means looking to our best scientific theories to determine what exists, or, perhaps more accurately, what we ought to believe to exist. Naturalism gives us some reason to believe in all such objects. Here is where the holism comes to the fore and, in particular, the confirmational holism. This position is the view that no claim of the theoretical science can be confirmed or refuted in isolation but only as a part of a system of hypotheses. So, if a theory is confirmed by empirical findings, the whole theory is confirmed. In particular, whatever mathematics is made use of in the theory is also confirmed. Furthermore, the same evidence that is appealed to in justifying belief in the mathematical components of the theory is appealed to in justifying the empirical portion of the theory.

One slightly different form of argument that deliver platonism as its conclusion is Resnik's *pragmatic indispensability argument* that focuses on the purpose of "doing science" and that has as an aim to avoid the reliance on confirmational holism that Quine's argument requires. The argument comes in two parts. The first is an argument for the conditional claim that if we are justified in drawing conclusions from and within science, then we are justified in taking mathematics used in science to be true. Resnik states this part of the argument as follows:

"1) In stating its laws and conducting its derivations science assumes the existence of many mathematical objects and the truth of much mathematics.

2) These assumptions are indispensable to the pursuit of science; moreover, many of the important conclusions drawn from and within science could not be drawn without taking mathematical claims to be true.

3) So we are justified in drawing conclusions from and within science only if we are justified in taking the mathematics used in science to be true."[43]

Resnik then combines the conclusion of this argument with the argument that we are justified in drawing conclusions from and within science, since this is the only way we know of doing science and that we are justified in doing science. The conclusion, then, is that we are justified in taking whatever mathematics is used in science to be true. The difference with Quine's argument is explained in this passage:

> "This argument is similar to the confirmational argument except that instead of claiming that the evidence for science (one body of statements) is also evidence for its mathematical components (another body of statements) it claims that the justification

[42] Quine, W. v. O., *Theories and Things*, p. 72.

[43] Resnik, M. D., "Scientific *vs*. Mathematical Realism: The Indispensability Argument," *Philosophia Mathematica*, v. 3, n. 2, (1995), pp. 169-170.

for doing science (one act) also justifies our accepting as true such mathematics as science uses (another act)."[44]

Notice that the argument does not require confirmation of any scientific theories in order for belief in mathematical objects to be justified. Indeed, even if all scientific theories were disconfirmed, we would presumably still need mathematics to do science, and since doing science is justified we would be justified in believing in mathematical objects. This is clearly a very powerful argument.

Finally, we must say that one consequence of the indispensability argument is that mathematical knowledge has an empirical character. Speaking of Quine this conclusion is quite normal. But it means that our knowledge of mathematical objects is *a posteriori*, and that it is plausible that mathematical knowledge is contingent. However, there are other questions to be answered: Does the indispensability argument yield acausal mathematical objects or does it tell us that mathematical objects are causally active after all? Does it tell us that we ought to believe in sets and sets alone, because mathematicians can build the rest of the mathematical universe out of them? It seems that the indispensability argument says nothing about all these issues. It simply asserts that there are mathematical objects. They might be constituted, for instance, by patterns or structures, as Resnik and Shapiro defend,[45] or as possible structures, as Putnam, Hellman and Parsons claim.[46] In short, any platonist account of mathematical objects is satisfactory by the indispensability argument. It must, of course, be consistent with the view that mathematics has an empirical character, which is the only real restriction.

The question of whether the indispensability argument delivers causal or acausal mathematical objects is a little more complicated. How could causally inert objects play an indispensable role in our best scientific theories? This question is answered by looking at the role such objects play in the relevant theories. The case here is no different from that of other theoretical objects. We do not conclude that electrons are causally active simply because they play an indispensable role in our theories of fundamental particles; we conclude that they are causally active because of the role they play in those theories. So too with mathematical objects. We must look at the role they play in our scientific theories. This role is, at least *prima facie*, not causal.

(2) Benacerraf's Indeterminacy Arguments[47]

In his "What Numbers Could Not Be," Benacerraf set out to undermine the prevailing view that numbers could be reduced to sets. This he did by pointing out that there is no unique reduction of the natural numbers to sets. For instance, we could have, according to

[44] RESNIK, M. D., "Scientific *vs*. Mathematical Realism: The Indispensability Argument," p. 171.

[45] Cf. RESNIK, M. D., *Mathematics as a Science of Patterns*, pp. 202-212 and 265-270; and SHAPIRO, S., *Philosophy of Mathematics. Structure and Ontology*, pp. 97-106 and 243-247.

[46] PUTNAM, H., "Mathematics Without Foundations," (1967), in PUTNAM, H., *Mathematics, Matter and Method*, 2nd. edition, pp. 43-59; HELLMAN, G., *Mathematics Without Numbers. Towards a Modal-Structural Interpretation*, Oxford University Press, Oxford, 1989; and PARSONS, C., "The Structuralist View of Mathematical Objects," (1990), in HART, W. D. (ed.), *The Philosophy of Mathematics*, Oxford University Press, Oxford, 1996, pp. 272-309.

[47] Benacerraf has revisited his arguments in "What Mathematical Truth Could Not Be-I," in MORTON, A. and STICH, S. P. (eds.), *Benacerraf and his Critics*, Blackwell, Oxford, 1996, pp. 9-59.

von Neumann, 3 = {Ø, {Ø}, {Ø, {Ø}}} or, according to Zermelo, 3 = {{{Ø}}}. Benacerraf then considers possible reasons for preferring one reduction over others, before deciding that there are no such reasons:

> "But if the number 3 is really one set rather than another, it must be possible to give some cogent reason for thinking so; for the position that this is an unknowable truth is hardly tenable. But there seems to be little to choose among the accounts."[48]

So he concludes that numbers could not be sets at all. Benacerraf's argument is directed primarily at reductionists who are not inclined to deny the existence of objects that have been successfully reduced to objects of another kind. That is, by reducing the natural numbers to some set-theoretical structure, all that is to be concluded is that we could do without the numbers since we have an isomorphic copy of them.

But Benacerraf pushes the point a little further and argues from two assumptions for the conclusion "that not only could numbers not be sets, they couldn't even be numbers:"[49] (1) the structuralist premise that arithmetic is the science of progressions (not a particular progression, but progressions in general), and (2) the claim that no (particular) system of objects could exhibit only structural properties. The thought is that arithmetic is concerned only with the structural properties of the natural numbers. Then Benacerraf considers what sorts of objects could have only structural properties and no intrinsic properties. He concludes that there can be no such objects because he believes that all legitimate objects must have some non-structural properties. However, it is clear that this argument has force only against structuralists. Indeed, numbers could be a variety of things, and the indispensability argument does not legislate against many of these in any way.

(3) Benacerraf's Dilemma

In his "Mathematical Truth," Benacerraf challenges philosophers of mathematics to: (1) naturalize mathematical epistemology, and (2) produce a semantics for mathematics that meshes with that of the rest of language. On a platonist account of mathematics the second challenge is met easily since a proposition such as "there exists a prime number greater than 3" is made true by the existence of the number 7 (among others), just as "there exists a city larger than Valencia" is made true by the existence of Barcelona (among others). The problem for platonism, however, is to provide a naturalized account of mathematical epistemology. Benacerraf also shows how various anti-platonist views of mathematical objects meet the first challenge but not the second. However, it is Benacerraf's challenge to platonism that his paper is best remembered for. And here is the whole argument:

Ontology and Tarski's theory of truth

(1) Any theory of mathematical truth must be in conformity with a general theory of truth. The best characterization of truth is Tarski's:

> "Its essential feature is to define truth in terms of reference (or satisfaction) on the basis of a particular kind of syntactico-semantical analysis of the language,

---

[48] Benacerraf, P., "What Numbers Could Not Be," p. 284.

[49] "What Mathematical Truth Could Not Be-I," p. 23.

and thus that any putative analysis of mathematical truth must be an analysis of a concept which is a truth concept at least in Tarski's sense."[50]

(2) Truth in mathematical propositions needs reference.

(3) That truth involves the existence of mathematical objects which are the referents of terms intertwined in those propositions.

(4) There must be infinitely many objects. So, they cannot be empirical or mental, but abstract, that is, non-spatio-temporal.

Epistemology and causal theory of knowledge

(1) In order to get knowledge there must be some connection between the knowing subject and the thing to be known:

"I favor a causal account of knowledge on which for *X* to know that *S* is true requires some causal relation to obtain between *X* and the referents of the names, predicates, and quantifiers of *S*. I believe in addition in a causal theory of *reference*, thus making the link to my saying knowingly that *S* is *doubly* causal."[51]

(2) If mathematical objects dwell in a non-spatio-temporal universe, they must be causally inert. So, there is no chance to get knowledge of those objects.

Conclusion

If we accept that mathematical objects exist as abstract objects and we accept the causal theory of knowledge, then *Benacerraf's dilemma* follows:

The thing required for truth in mathematics make impossible the knowledge of that truth; the thing that would make possible mathematical knowledge makes impossible the truth of that knowledge.

Therefore, if platonism is true, mathematical knowledge is impossible; and if it is supposed that we have mathematical knowledge, then platonism must be false.

It is sufficient to note that Benacerraf's challenge to platonism explicitly depends on a now largely discredited epistemology, the causal theory of knowledge.[52] But Field suggests that Benacerraf's problem for that position may be restated as a problem of explaining the reliability of our mathematical beliefs:

"Benacerraf's challenge —or at least, the challenge which his paper suggests to me— is to provide an account of the mechanisms that explain how our beliefs about these remote entities can so well reflect the facts about them. The idea is that if *it appears in principle impossible to explain this*, then that tends to *undermine* the belief

[50] Benacerraf, P., "Mathematical Truth," p. 408.

[51] "Mathematical Truth," p. 412.

[52] A. Goldman devised this theory in 1967. At the beginning, he put forward that, in order to be justified, a belief must be caused by the fact that made it true. But this view could not held as a sufficient condition for non-inferential perceptual beliefs. Moreover, it left up in the air the knowledge about the future —even mathematical knowledge!— and it could not explain how a fact by itself was able to lead to beliefs. For these reasons, Goldman and other scholars modified it introducing the requirement that, in order to be justified, the belief must be obtained by a reliable method or through some other virtuous devises involved in the formation of beliefs. Cf. Goldman's papers in Pappas, G. S. and Swain, M. (eds.), *Essays on Knowledge and Justification*, Cornell University Press, Ithaca, NY, 1978.

in mathematical entities, *despite* whatever reasons we might have for believing in them. Of course, the reasons for believing in mathematical entities (in particular, the indispensability arguments) still need to be addressed, but the role of the Benacerrafian challenge (as I see it) is to raise the cost of thinking that the postulation of mathematical entities is a proper solution, and thereby increase the motivation for showing that mathematics is not really indispensable after all."[53]

This formulation of Benacerraf's problem does not depend on the causal theory of knowledge. How might an indispensabilist respond to this challenge? The short answer is that the indispensability argument tells us that we come by mathematical knowledge in exactly the same way as other forms of knowledge, i.e., by the hypothetico-deductive methods of science. Let us note that this is a question that platonists and nominalists alike must answer. And Field's answer depends on mathematics being contingent:

"One could argue, for instance, that if mathematics is indispensable to laws of empirical science, then *if the mathematical facts were different, different empirical consequences could be derived from the same laws of (mathematized) physics*. So, it could be said mathematical facts make an empirical difference, and maybe this would enable the application-based platonist to argue that our observations of the empirical consequences of physical law are enough to explain the reliability of our mathematical beliefs."[54]

So, as we said, this is the most plausible position for an indispensabilist to take in any case. In addition, the reliability of our mathematical beliefs may be explained by their applicability to other areas of mathematics, so long as there is a chain of applications that leads eventually to empirical evidence.[55] Such is the nature of applicability in the Quinean holistic view of science. But the Benacerraf challenge asks for justification of our mathematical beliefs *one at a time*. According to the indispensability argument, though, mathematical beliefs —indeed, all beliefs— are justified holistically. Someone who subscribes to this argument believes in mathematical objects because of the role they play in our total theory. Such a person should feel no compulsion to justify mathematical beliefs individually, because in this framework it could be entirely unreasonable.

Incidentally, Putnam has shown some worries about the obscurity of the notion of "naturalizing epistemology" in the area of philosophy of (logic and) mathematics. He has pointed out that the trouble is that many of our key notions ("understanding something", "something's being capable of being confirmed, or discovered to be true", etc.) are normative notions. In the case of mathematics, the trouble with epistemology "depends on the idea that there is a problem of *justification* in this area. But —Putnam keeps on saying— perhaps mathematics does not require justification; it only require theorems."[56] Surely Putnam knows that theorems require proof, a mathematical and epistemic notion,

---

53 Field, H., *Realism, Mathematics, and Modality*, p. 26.

54 *Realism, Mathematics, and Modality*, pp. 28-29.

55 But not only applications. Mathematicians may see some other connections: "I strongly disagree with the view that mathematics is simply a collection of separate subjects, that you can invent a new branch of mathematics by writing down axioms 1, 2, 3 and going away and working on your own. Mathematics is much more of an organic development. It has a long history of connections with the past and connections with other subjects," Atiyah, M., "An Interview with Michael Atiyah," *The Mathematical Intelligencer*, v. 6, n. 1, (1984), p. 11.

56 Putnam, H., *Words and Life*, Harvard University Press, Cambridge, MA, 1994, p. 260.

and a tool for mathematicians, because it is a convenient and natural way to express the fact that the assertion of a mathematical statement is internally justified *only by a proof of* it. Of course, this notion of justification is open-ended; but that is simply an expression of the essential incompleteness of mathematics; it is not a further problem arising from it. And externally, can anyone imagine a better way to justify mathematics than the applications to science? Even once Putnam wrote that "mathematical knowledge resembles *empirical* knowledge —that is, that the criterion of truth in mathematics just as much as in physics is success of our ideas in practice, and that mathematical knowledge is corrigible and not absolute."[57]

Finally, we must admit that the question of whether Quinean indispensabilists have a satisfactory answer to Benacerraf's challenge is a difficult and complicated matter. However, we hope to have shown that from a Quinean point of view there is still a possible answer to the challenge. Anyway, about the choice between platonism and anti-platonism, as Balaguer says "there is no fact of the matter as to which of them is true." Perhaps, borrowing some words from Lakatos, "the answer will scarcely be a monolithic one."[58]

3. Objectivity in Mathematics

We have already noticed that it is possible to be platonist without being committed to mathematical objects: it is enough to be some sort of objectivist. The dominant use of "objectivity" these days belongs to epistemology or methodology. Methodological objectivity is primarily a feature of inquiries or methods of inquiry, and derivatively of the people who conduct inquiries and the judgments they form as a result. To a first approximation, we call an inquiry "objective" when its trajectory is unaffected in relevant ways by the peculiar biases, preferences, ideological commitments and the like of the people who conduct it. But it seems that, according to the rhetoric that surrounds this topic, one of the hallmarks of the objective is precisely its (in)dependence of our theories or conceptions. We may speak of objective properties and relations, and in some cases even of objective objects. For instance, P is an objective property if it is an objective fact that an object possesses it; and the object x is objective if the fact that x exists is an objective fact.

Given this way of speaking, the question of the objectivity of mathematics is to be understood as a question about whether properties like truth are features of some objective mathematical world. In other words, are the mathematical facts entirely independent of our mathematical thinking, or are they rather somehow constructed by it? The problem is one that can be thought to be intensified without ontology, in particular without real or abstract objects for our language to refer to. On the other hand, real or abstract objects to which the language is bound can be thought to make one's or everyone's reasoning irrelevant or at least to play second fiddle to the previously existing relationships to which one does not have access. A lack of pre-existing objects, or ignoring them if they should happen to exist, allows us the freedom to use various kind of logic instead of having to depend on the exactly right one, whichever it is. Often referred to in this

[57] Putnam, H., "What is Mathematical Truth?," p. 61.

[58] Lakatos, I., *Mathematics, Science and Epistemology*, Cambridge University Press, Cambridge, 1978, p. 41.

connection is Kreisel.[59] Many mathematicians will subscribe to the old fashioned idea, as expressed by him, that the analysis of intuitive notions or concepts gives lace to rules and definitions that fix their properties. But then what "the "old fashioned" idea assumes is quite simply that the intuitive notions are significant, be it in the external world or in thought (and a precise formulation of what is significant in a subject is the result, not a starting point, of research into that subject)." But the significant notions are relational,[60] at bottom as objective and sharable as objects. So, Kreisel adds:

> "Informal rigour[61] wants (i) to make this analysis as precise as possible (…) in particular to eliminate doubtful properties of the intuitive notions when drawing conclusions about them, and (ii) to extend their analysis, in particular, not to leave undecided questions which can be decided by full use of evident properties of these intuitive notions."[62]

Kreisel notes the difficulty of the mathematical concepts to be analyzed and, considering the lack of precision, the necessity of recognizing the objectivity of some basic notions. Speaking about set theory, these notions are "subset" and "power set". However, it is involved the contrast "objective vs. subjective" that Kreisel compare with the more familiar contrast "heredity vs. environment."

The second element of each pair is problematic and sensitive to "accidental circumstances" and "lethal mutations," resulting in "minor variants." It must mean that for different reasons mathematical concepts are suffering some change or some kind of "evolution."[63] That is, that "mathematical concepts are not fixed once and for all," and that mathematicians are trying to sharpen their understanding of them as well as possible, through "reflective expansion."[64] In fact, Kreisel points out to mathematical practice where, "having reflected carefully, mathematicians find their reasoning [with those abstract concepts] full convincing," and their elimination would add nothing "to their conviction."[65]

---

59 Cf. Kreisel, G., "Wittgenstein's Remarks on the Foundations of Mathematics," *The British Journal for the Philosophy of Science*, v. 9, (1958-59), pp. 135-158; and Kreisel, G., "Informal Rigour and Completeness Proofs," in Lakatos, I. (ed.), *Problems in the Philosophy of Mathematics*, North-Holland, Amsterdam, 1967, pp. 138-171.

60 As Frege said, "a concept can be fixed only by its relations to other concepts. These relations, formulated in certain statements, I call axioms," Frege, G., *Philosophical and Mathematical Correspondence*, edited by G. Gabriel et al., Blackwell, Oxford, 1980, p. 51.

The mathematician I. Stewart claims that in mathematics we studies the ways in which different concepts are related to each other. Cf. Stewart, I., *From Here to Infinity. A Guide to Today's Mathematics*, Oxford University Press, Oxford, 2nd. revised edition, 1996. Translated into Spanish: Stewart, I., *De aquí al infinito. Las Matemáticas hoy*, Crítica, Barcelona, 1998, p. 14.

61 Here we have how Kreisel explains the meaning of 'informal rigour':

"Informal rigour is involved in the 2000-year-old tradition of analysing common notions in mathematical terms (current at a given time). In particular, definitions and other properties of such notions are established in rigorous ways. When this way consists of inspection by the mind's eye, one speaks of 'axioms' in the old sense of this word, now called 'informal'. When this way consists of formal deduction from established knowledge, one speaks of formal rigour. Taken literally, the domain of notions that are 'common' varies in space and time (and other dimensions)," Kreisel, G., "Church's Thesis and the Ideal of Informal Rigour," *Notre Dame Journal of Formal Logic*, v. 28, n. 4, (1987), p. 502.

62 Kreisel, G., "Informal Rigour and Completeness Proofs," pp. 138-139.

63 See the final lines in Kreisel's quote included in the footnote n. 61.

64 Feferman, S., *In the Light of Logic*, Oxford University Press, Oxford, 1998, p. 120.

65 Kreisel, G., "Two Notes on the Foundations of Set-Theory," *Dialectica*, v. 23, n. 2, (1969), p. 96.

However, the formulation of mathematical principles, which evidently makes always for precision, encapsulates the properties of concepts before reaching some conclusions about them. Interestingly enough, Kreisel states that "there is no claim that these particular properties describe exhaustively our understanding of the, possibly abstract, notion considered."[66] In addition, after asking himself "Is mathematics about an external reality, about objects which we conceive as being independent of us?", Kreisel notices that the consideration of the objectivity of some concepts may be pursued without answering whether some reality external to ourselves is involved, because the issues considered do not depend on it. It happens that the notions "come to us naturally and not as the result of some deliberate choice, just because "we cannot help ourselves" as one says."[67] Moreover, it would be difficult to distinguish them "from something given to us from outside; even if clever analysis were to show that they aren't."[68]

But, the relation between objectivity and that external reality, "though unconvincing, is probably not particularly misleading: if something in us is not under control, if it is part of the data, to a first approximation it will be considered as "externally" given."[69] So to say that a mathematical statement exhibits objectivity is to say that it may be intelligible to a mathematician even though resolving its truth-value may defeat our cognitive powers. For instance, take Goldbach's Conjecture. We may understand its statement, but we are in principle incapable of resolving its truth-value. This gives us one way of claming that mathematical claims are more objective than non-mathematical claims.

We feel that Quine's moderated platonism and the kind of platonism, close to Frege's objective content of thought, professed by Popper may help us to complete Kreisel's conception of objectivity. Our beliefs are constrained by considerations of plausibility, coherence with firmly prior doctrine, elegance, systematicity, and the like; but not by anything like an encounter with the objects or their visible traces.[70] These are Quine's "considerations of equilibrium" that affect the field as a whole[71], or as explained in one of his last papers, where sentences "vie with one another in a surging equilibrium of evidential claims. Such (...) is the web of belief."[72] Then the proverbial necessity of mathematical truth resides merely in our exempting the mathematical sentences when choosing which one of a refuted block of sentences to revoke. We exempt them because changing them would reverberate excessively through science.

There is no deeper sense of "reality" than the sense in which it is the business of science itself, by its self-corrective hypothetico-deductive method of conceptualization and experiment, to seek the essence of reality. Implication, as defined by elementary

---

[66] Kreisel, G., "The Formalist-Positivist Doctrine of Mathematical Precision in the Light of Experience," *L'Âge de la Science*, v. 3, n. 1, (1970), p. 21.

[67] This "one" could be Gödel. As reported later on, for him axioms "force themselves upon us as being true," Gödel, K., "What is Cantor's Continuum Problem?," p. 268.

[68] Kreisel, G., "The Formalist-Positivist Doctrine of Mathematical Precision in the Light of Experience," p. 20.

[69] Kreisel, G., "Two Notes on the Foundations of Set-Theory," p. 97.

[70] Notice that Popper ("On the Theory of the Objective Mind") seems to hold that mathematical objects can causally affect our immaterial minds.

[71] "Two Dogmas of Empiricism," p. 43.

[72] Quine, W. v. O., "I, You, and It, An Epistemological Triangle," in Orenstein A. and Kotatko, P. (eds.), *Knowledge, Language and Logic, Questions for Quine*, Kluwer, Dordrecht, 2000, p. 5.

predicate logic, is the lifeblood of theories, and what relates a theory to its checkpoints in observation categoricals.[73] Observation sentences are "the vehicle of evidence for objective science."[74] Here Quine is also close to Popper's analogy of science as "an edifice supported only by a multitude of long piles driven deep down into a bottomless swamp[75]."[76] Even the concept of a system whose elements must stand in mutual relations of consistency and interdependence, and in openness to revision is present in Popper's thought: "The demand for scientific objectivity makes it inevitable that every scientific statement must remain *tentative for ever*. It may indeed be corroborated, but every corroboration is relative to other statements which, again, are tentative."[77]

It is true that the main mathematical activity consists of the investigation of the consequences of the accepted axioms. But before the mathematician is supposed to determine the intuitive nature of the studied objects. For this matter, s/he makes use of a network of concepts and operations applicable to those objects. S/he is supposed to present also the main propositions or axioms that will define implicitly those concepts and operations, and that eventually will require some adjustment.[78] The reasons that are accustomed to have the mathematicians in order to introduce axioms reside partly in that they allow to restrict arbitrary definitions of the implied concepts and in that they limit the validity of the claims about them.

Nevertheless, the axioms condense a series of ideas that turn up in the process of search that ends in their formulation. In order to avoid infinite regress in mathematical justification in an strict sense (i.e., proof), we are also lead to adopt the axioms as starting points. This is an aspect of the question that shows to what extent it is convenient to present reasons to favor the axioms, because it is not that these are introduced in an arbitrary way or that reasons are definitive, and because they will always remain open to debate and will be analyzed critically, so that we remain convinced for the better ones.[79] That is the reason why Popper made a difference between justification and criticism of theories and beliefs.[80]

Now then, upon talking about the mathematical truth, we are talking about the "correspondence with the mathematical facts" and something more, something that does

---

[73] QUINE, W. v. O., *From Stimulus to Science*, Harvard University Press, Cambridge, MA, 1995, 51.

[74] QUINE, W. v. O., "In Praise of Observation Sentences," *Journal of Philosophy*, 90, n. 3, (1993), p. 109.

[75] Cf. POPPER, K. R., *The Logic of Scientific Discovery*, Hutchinson, London, 1959; 9th. impression 1977, p. 111: "The empirical basis of objective science has thus nothing 'absolute' about it. Science does not rest upon solid bedrock. The bold structure of its theories rises, as it were, above a swamp. It is like a building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or 'given' base; and if we stop driving the piles deeper, it is not because we have reached firm ground. We simply stop when we are satisfied that the piles are firm enough to carry the structure, at least for the time being."

[76] QUINE, W. v. O., "Response to Lehrer," in ORENSTEIN A. and KOTATKO, P. (eds.), *Knowledge, Language and Logic, Questions for Quine*, p. 412.

[77] POPPER, K. R., *The Logic of Scientific Discovery*, p. 280.

[78] This appreciation finds inspiration in HILBERT, D., "Axiomatisches Denken," (1918), reprinted in HILBERT, D., *Gesammelte Abhandlungen*, Chelsea, Bronx, NY, 1965, vol. 3, pp. 146-156.

[79] Cf. MADDY, P., *Naturalism in Mathematics*, Clarendon Press, Oxford, 1997; and ALCOLEA, J., LORENTE, J. M. and ÚBEDA, J. P., "Justificación y racionalidad de axiomas," in NEPOMUCENO A. ET AL. (eds.), *Logic, Language and Information*, Editorial Kronos, Sevilla, 2000, pp. 1-7.

[80] Cf. POPPER, K. R., *Realism and the Aim of Science. From the Postscript to the Logic of Scientific Discovery: Vol. I*, edited by W. W. Bartley III, Hutchinson, London, 1983; reprinted by Routledge, London, 1992, pp. 28-29.

not reside in us, since the relation of correspondence could be the case without our knowing it. These considerations took Popper to a theory of objective knowledge, of the knowledge that exists outside of us and, in part, independently of us. This theory takes the form of a pluralistic metaphysics of the well-known three worlds. The necessity of the third world is given expression in the necessity of separating the very argument of the beliefs and desires of the participants in the dialogue, because it is necessary to show the rationality of the process and it could only be done identifying after the successive improvement the underlying argument, that lives and is developed in the third world together with propositions, theories, problems, etc.

Notice that the requirement of intersubjectivity is what makes then mathematics objective.[81] And, as Popper says, "the objectivity, even of intuitionist mathematics, rests, as does that of all science, upon the criticizability of its arguments. But this means that language becomes indispensable as the medium of argument, of critical discussion."[82]

The analysis of the process of critical discussion of conjectures makes clear that the discussion never considers the question if a conjecture is "justified" in the sense that *we* are justified to accept it as *true*. At the very best, the critical discussion justifies the statement that the conjecture in question is the best that it is provided, or, in other words, that it is the one that comes nearer to the truth. The process of critical discussion is a process of rational discussion. The participants are willing to discuss critically their own beliefs and to correct them in the light of the critical discussions with other people. It is a principle of minimum rationality the one that encourages our conjectures and our proofs, and although we know that our conjecture is not true, we have some reason in order to consider it as a good approximation to the truth. So, the state of the mathematics is unavoidably hypothetical, and the mathematics of the moment describes the mathematical reality in an approximate way. There is no reason to suppose that that description is trustworthy. The thesis that the mathematics describes an objective mathematical world should be considered more as a question of intention than as an accomplished fact. Beyond the axioms that we have accepted as properly analyzing our mathematical concepts, there remain, and always will, statements which are undecided by them. So, mathematics becomes speculative in the sense that even the most elementary deductions or statements must answer to a reality which we, at best, can only partially comprehend and about which we could be wrong.

We are not in conditions of affirming that mathematics, as it is, reaches a characterization of the mathematical reality. In spite of everything, the mathematicians continue having

---

[81] "The requirement of intersubjectivity is what makes science objective," Quine, W. v. O., *Pursuit of Truth*, Harvard University Press, Cambridge, MA, 1990, pp. 4-5.

Compare with Poincaré: "What is objective must be common to many minds and consequently transmissible from one to the other. (...) Nothing, therefore, will have objective value except what is transmissible by 'discourse', that is, intelligible," Poincaré, H., *The Foundations of Science*, The Science Press, New York, 1913, p. 347 and 349).

It is also instructive to compare Quine and Popper's position with Feferman's: thinking that platonism is "unsatisfactory" for justifying set theory and mathematics, Feferman says that "some other philosophy grounded in inter-subjective *human* conceptions will have to be sought to explain the apparent objectivity of mathematics," Feferman, S., "Does Mathematics Need New Axioms?," *The American Mathematical Monthly*, v. 106, n. 2 (1999), p. 110.

[82] Popper, K. R., "Epistemology Without a Knowing Subject," pp. 136-137.

realistic leanings. So, to this respect, Gödel's position was very important. He claimed *clarity* to the general mathematical concepts "for us to be able to recognize their soundness and the truth of the axioms concerning them," which sometimes "force themselves upon us as being true." This is a sign of the action of the third world on the second world. Now, how do we obtain that clarity and correction, if not upon reflection? The consequence would be that "the set-theoretical concepts and theorems describe some well-determined reality," although the system would always be incomplete and would need of the addition of new axioms.[83] The *aim* of mathematics is, without a doubt, to answer to our questions about the mathematical world *correctly* and to describe that world "as it is in reality." For Gödel, Quine and Popper, as for Kreisel, the orientation of the mathematics is objective, but also factual, because it takes care of establishing *true* facts on the real mathematical world, a world that surpasses our cognitive grasp.

4. Bibliography

Alcolea, J., "Un aspecto del intuicionismo matemático," *Quaderns de Filosofia i Ciència*, n. 9-10, (1986), pp. 13-20.

Alcolea, J., "Arend Heyting," *Mathesis*, v. 4, n. 2, (1988), pp. 189-220.

Alcolea, J., Lorente, J. M. and Ubeda, J. P., "Justificación y racionalidad de axiomas," in Nepomuceno A. et al. (eds.), *Logic, Language and Information*, Editorial Kronos, Sevilla, 2000, pp. 1-7.

Atiyah, M., "An Interview with Michael Atiyah," *The Mathematical Intelligencer*, 6, n. 1, (1984), pp. 9-19.

Balaguer, M., *Platonism and Anti-Platonism in Mathematics*, Oxford University Press, Oxford, 1998.

Barrett, W., *Irracional Man. A Study in Existential Philosophy*, Doubleday Anchor Books, Garden City, NY, 1962.

Benacerraf, P., "What Numbers Could Not Be," *Philosophical Review*, v. 74, (1965), pp. 47-73. Reprinted in Benacerraf, P. and Putnam, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, Cambridge University Press, Cambridge, 1983, pp. 272-294.

Benacerraf, P., "Mathematical Truth," *Journal of Philosophy*, v. 70, (1973), pp. 661-680. Reprinted in Benacerraf, P. and Putnam, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, Cambridge University Press, Cambridge, 1983, pp. 403-420.

Benacerraf, P., "What Mathematical Truth Could Not Be-I," in Morton, A. and Stich, S. P. (eds.), *Benacerraf and his Critics*, Blackwell, Oxford, 1996, pp. 9-59.

Bernays, P., "Sur le platonisme dans les mathématiques," *L'Enseignement Mathématique*, v. 34, (1935). Translated into English by C. D. Parsons: "On Platonism in Mathematics," in Benacerraf, P. and Putnam, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, Cambridge University Press, Cambridge, 1983, pp. 258-271.

Bigelow, J., *The Reality of Numbers. A Physicalist's Philosophy of Mathematics*, Clarendon Press, Oxford, 1988.

[83] Gödel, K., "What is Cantor's Continuum Problem?," pp. 258, 260 and 268.

BROUWER, L. E. J. (1949), "Consciousness, Philosophy, and Mathematics," in BETH, E. W., POS, H. J., and HOLLAK, J. H. A. (eds.), *Proceedings of the Tenth International Congress of Philosophy*, North Holland, Amsterdam, 1949, vol. 1, pp. 1235-1249. Reprinted in BENACERRAF, P. AND PUTNAM, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, Cambridge University Press, Cambridge, 1983, pp. 90-96.

BROUWER, L. E. J., *Collected Works.* Vol. 1. *Philosophy and Foundations of Mathematics*, edited by A. Heyting, North-Holland, Amsterdam, 1975.

BURGESS, J. and ROSEN, G., *A Subject with No Object. Strategies for Nominalistic Interpretation of Mathematics,* Clarendon Press, Oxford, 1997.

CARNAP, R., "Die logizistische Grunlegung der Mathematik", *Erkenntnis*, v. 2, (1931), pp. 91-121. Translated into English by H. Putnam and G. J. Massey: "The Logicist Foundations of Mathematics," in BENACERRAF, P. and PUTNAM, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, Cambridge University Press, Cambridge, 1983, pp. 41-51.

CARNAP, R., "Empiricism, Semantics and Ontology," *Revue International de Philosophie*, v. 4, (1950), pp. 20-40. Reprinted in BENACERRAF, P. and PUTNAM, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, Cambridge University Press, Cambridge, 1983, pp. 241-257.

DUMMETT, M., "Realism," Oxford University Philosophical Society, 8. 3. 1963. Reprinted in DUMMETT, M., *Truth and Other Enigmas*, Duckworth, London, 1978, pp. 145-165.

DUMMETT, M., "The Philosophical Basis of Intuitionistic Logic", en ROSE, H. E. and SHEPHERDSON, J. C. (eds.), *Logic Colloquium '73*, North Holland, Amsterdam, 1975, pp. 5-40. Reprinted in BENACERRAF, P. and PUTNAM, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, Cambridge University Press, Cambridge, 1983, pp. 97-129.

FEFERMAN, S., *In the Light of Logic,* Oxford University Press, Oxford, 1998.

FEFERMAN, S., "Does Mathematics Need New Axioms?," *The American Mathematical Monthly,* v. 106, n. 2, (1999), pp. 99-111.

FIELD, H., *Science Without Numbers. A Defence of Nominalism,* Blackwell, Oxford, 1980.

FIELD, H., *Realism, Mathematics, and Modality,* Blackwell, Oxford, 1989.

FREGE, G., *Die Grundlagen der Arithmetik. Eine logischmathematische Untersuchung über den Begriff der Zahl,* Koebner, Breslau, 1884. Translated into English by J. L. Austin: *The Foundations of Arithmetic,* Blackwell, Oxford, 2nd. revised edition 1953.

FREGE, G., *Wissenschaftlicher Briefwechsel,* edited by G. Gabriel, H. Hermes, F. Kambartel, C. Thiel and A. Veaart, F. Meiner, Hamburg, 1976. Translated into English by H. Kaal and edited by B. McGuinness: *Philosophical and Mathematical Correspondence,* B. Blackwell, Oxford, 1980.

FREGE, G., *Collected Papers on Mathematics, Logic, and Philosophy*, edited by B. McGuinness et al., Blackwell, Oxford, 1984.

GÖDEL, K., "Some Basic Theorems on the Foundations of Mathematics and their Philosophical Implications," (1951), in GÖDEL, K., *Collected Works.* Vol. III. *Unpublished Essays and Lectures*, edited by S. Feferman et al., Oxford University Press, Oxford, 1995, pp. 304-323.

GÖDEL, K., "What is Cantor's Continuum Problem?," in BENACERRAF, P. and PUTNAM, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, Prentice-Hall, Englewood-Cliffs, NJ, 1964, pp. 258-273. Also in BENACERRAF, P. and PUTNAM, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, Cambridge University Press, Cambridge, 1983, pp. 470-485. Reprinted in GÖDEL, K., *Collected Works. Vol. II. Publications 1938-1974*, edited by S. Feferman et al., Oxford University Press, Oxford, 1990, pp. 254-270.

GONZALEZ, W. J., "'Verdad' y 'prueba' ante el problema del progreso matemático," in MARTINEZ FREIRE, P. F. (ed.), *Filosofía actual de la Ciencia*, Publicaciones Universidad de Málaga, Málaga, 1998, pp. 307-346.

HARDY, G. H.*, A Mathematician's Apology*, Cambridge University Press, London, 1940. Translated into Spanish: HARDY, G. H., *Autojustificación de un matemático*, Ariel, Barcelona, 1981.

HELLMAN, G., *Mathematics Without Numbers. Towards a Modal-Structural Interpretation*, Oxford University Press, Oxford, 1989.

HEYTING, A., "Die intuitionistische Grundlegung der Mathematik", *Erkenntnis*, v. 2, (1931), pp. 106-115. Translated into English by H. Putnam and G. J. Massey: "The Intuitionist Foundations of Mathematics," in BENACERRAF, P. and PUTNAM, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, Cambridge University Press, Cambridge, 1983, pp. 52-60.

HEYTING, A. "After Thirty Years," in NAGEL, E., SUPPES, P. and TARSKI, A. (eds.), *Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress*, Stanford University Press, Stanford, CA, 1962, pp. 194-197.

HILBERT, D., "Axiomatisches Denken," *Mathematische Annalen,* v. 78, (1918), pp. 405-415. Reprinted in HILBERT, D., *Gesammelte Abhandlungen,* Chelsea, Bronx, NY, 1965, vol. 3, pp. 146-156.

HILBERT, D., "Über das Unendliche," *Mathematische Annalen,* v. 95, (1926), pp. 161-190. English translation: "On the Infinite," (1925), reprinted in BENACERRAF, P. and PUTNAM, H. (eds.), *The Philosophy of Mathematics. Selected Readings,* Cambridge University Press, Cambridge, 1983, pp. 183-201.

KREISEL, G., "Wittgenstein's Remarks on the Foundations of Mathematics," *British Journal for the Philosophy of Science*, v. 9, (1958-59), pp. 135-158.

KREISEL, G., "Informal Rigour and Completeness Proofs," in LAKATOS, I. (ed.), *Problems in the Philosophy of Mathematics*, North-Holland, Amsterdam, 1967, pp. 138-171.

KREISEL, G., "Two Notes on the Foundations of Set-Theory," *Dialectica*, v. 23, n. 2, (1969), pp. 93-114.

KREISEL, G., "The Formalist-Positivist Doctrine of Mathematical Precision in the Light of Experience," *L'Âge de la Science*, v. 3, n. 1, (1970), pp. 17-46.

KREISEL, G., "Church's Thesis and the Ideal of Informal Rigour," *Notre Dame Journal of Formal Logic*, v. 28, n. 4, (1987), pp. 499-519.

LAKATOS, I., *Mathematics, Science and Epistemology*, Cambridge University Press, Cambridge, 1978.

MADDY, P., *Realism in Mathematics*, Clarendon Press, Oxford, 1990.

MADDY, P., *Naturalism in Mathematics*, Clarendon Press, Oxford, 1997.

MYHILL, J., "Remarks on Continuity and the Thinking Subject," in LAKATOS, I. (ed.), *Problems in the Philosophy of Mathematics*, North-Holland, Amsterdam, 1967, pp. 173-175.

PAPPAS, G. S. and SWAIN, M. (eds.), *Essays on Knowledge and Justification*, Cornell University Press, Ithaca, NY, 1978.

PARSONS, C., "The Structuralist View of Mathematical Objects," *Synthese*, v. 84, (1990), pp. 303-346. Reprinted in HART, W. D. (ed.), *The Philosophy of Mathematics*, Oxford University Press, Oxford, 1996, pp. 272-309.

POINCARÉ, H., *The Foundations of Science*, The Science Press, New York, 1913.

POPPER, K. R., "The Aim of Science," *Ratio*, v. 1, n. 1, (1957), pp. 24-35. Reprinted in POPPER, K. R., *Objective Knowledge*, Clarendon Press, Oxford, 1972, pp. 191-205.

POPPER, K. R., *The Logic of Scientific Discovery*, Hutchinson, London, 1959; 9th impression 1977.

POPPER, K. R., "Epistemology Without a Knowing Subject," in VAN ROOTSELAAR, B. and STAAL, J. F. (eds.), *Logic, Methodology and Philosophy of Science III. Proceedings of the Third International Congress for Logic, Methodology and Philosophy of Science*, North-Holland, Amsterdam, 1968, pp. 333-373. Reprinted in POPPER, K. R., *Objective Knowledge*, Clarendon Press, Oxford, 1972, pp. 106-152.

POPPER, K. R., "On the Theory of the Objective Mind," in POPPER, K. R., *Objective Knowledge. An Evolutionary Approach*, Clarendon Press, Oxford, 1972, pp. 153-190.

POPPER, K. R., *Realism and the Aim of Science. From the Postscript to the Logic of Scientific Discovery: Vol. I*, edited by W. W. Bartley III, Hutchinson, London, 1983; reprinted by Routledge, London, 1992.

PUTNAM, H., "Mathematics without Foundations", *Journal of Philosophy,* v. 64, (1967), pp. 5-22. Reprinted in PUTNAM, H., *Mathematics, Matter and Method*, Cambridge University Press, Cambridge, MA, 2nd edition, 1979, pp. 43-59.

PUTNAM, H., *Philosophy of Logic*, G. Allen and Unwin, London, 1971. Partially reprinted in PUTNAM, H., *Mathematics, Matter and Method*, Cambridge University Press, Cambridge, MA, 2nd. edition, 1979, pp. 323-357.

PUTNAM, H. "What is Mathematical Truth?," *Historia Mathematica*, v. 2, (1975), pp. 529-543. Reprinted in PUTNAM, H., *Mathematics, Matter and Method*, Cambridge University Press, Cambridge, MA, 2nd. edition, 1979, pp. 60-78.

PUTNAM, H., *Words and Life*, Harvard University Press, Cambridge, MA, 1994.

QUINE, W. v. O., "On What There is," *Review of Metaphysics*, v. 2, (1948), pp. 21-38. Reprinted in QUINE, W. v. O., *From a Logical Point of View*, Harper and Row, New York, 2nd. edition revised, 1961, pp. 1-19.

QUINE, W. v. O., "Two Dogmas of Empiricism," *Philosophical Review*, v. 60, (1951), pp. 20-46. Reprinted in QUINE, W. v. O., *From a Logical Point of View*, Harper and Row, New York, 2nd. edition revised, 1961, pp. 20-46.

QUINE, W. v. O., "The Scope and Language of Science," (1957), in QUINE, W. v. O., *The Ways of Paradox and Other Essays*, Cambridge University Press, Cambridge, MA, revised and enlarged edition, 1976, pp. 228-245.

QUINE, W. v. O., "Epistemology Naturalized," in QUINE, W. v. O., *Ontological Relativity and Other Essays*, Columbia University Press, New York, 1969, pp. 69-90.

QUINE, W. v. O., *Theories and Things*, Harvard University Press, Cambridge, MA, 1982.

QUINE, W. v. O., *Pursuit of Truth*, Harvard University Press, Cambridge, MA, 1990.

QUINE, W. v. O., "Two Dogmas in Retrospect," *Canadian Journal of Philosophy*, v. 21, n. 3, (1991), pp. 265-274.

QUINE, W. v. O., "In Praise of Observation Sentences," *Journal of Philosophy*, v. 90, n. 3, (1993), pp. 107-116.

QUINE, W. v. O., *From Stimulus to Science*, Harvard University Press, Cambridge, MA, 1995.

QUINE, W. v. O., "I, You, and It, An Epistemological Triangle," in ORENSTEIN A. and KOTATKO, P. (eds.), *Knowledge, Language and Logic, Questions for Quine*, Kluwer, Dordrecht, 2000, pp. 1-6.

QUINE, W. v. O., "Response to Lehrer," in ORENSTEIN A. and KOTATKO, P. (eds.), *Knowledge, Language and Logic, Questions for Quine*, Kluwer, Dordrecht, 2000, p. 412.

RESNIK, M. D., "Scientific *vs.* Mathematical Realism: The Indispensability Argument," *Philosophia Mathematica*, v. 3, n. 2, (1995), pp. 166-174.

RESNIK, M. D., *Mathematics as a Science of Patterns*, Clarendon Press, Oxford, 1997.

SHAPIRO, S., *Philosophy of Mathematics. Structure and Ontology*, Oxford University Press, Oxford, 1997.

STEWART, I., *From Here to Infinity. A Guide to Today's Mathematics*, Oxford University Press, Oxford, 2nd. revised edition, 1996. Translated into Spanish: STEWART, I., *De aquí al infinito. Las Matemáticas hoy*, Crítica, Barcelona, 1998.

TYMOCZKO, T., "Mathematics, Science and Ontology," *Synthese*, v. 88, (1991), pp. 201-228.

TROELSTRA, A. S. and VAN DALEN, D., *Constructivism in Mathematics. An Introduction*. North-Holland, Amsterdam, vol. 1, 1988.

VON NEUMANN, J., "Die formalistische Grundlegung der Mathematik", *Erkenntnis*, v. 2, (1931). Translated into English by H. Putnam and G. J. Massey: "The Formalist Foundations of Mathematics," in BENACERRAF, P. and PUTNAM, H. (eds.), *The Philosophy of Mathematics. Selected Readings*, Cambridge University Press, Cambridge, 1983, pp. 61-65.

Subject Index

Index of Names