Francesco Guala
London School of Economics and Political Science

Economics and the Laboratory

Some Philosophical and Methodological Problems Facing Experimental Economics

In partial fulfillment of the degree of Philosophy Doctor

1999

Abstract:

Laboratory experimentation was once considered impossible or irrelevant in economics. Recently, however, economic science has gone through a real 'laboratory revolution', and experimental economics is now a most lively subfield of the discipline. The methodological advantages and disadvantages of controlled experimentation constitute the main subject of this thesis. After a survey of the literature on experiments in philosophy and economics (chapter one), the problem of testing normative theories of rationality is tackled (chapter two). This philosophical issue was at the centre of a famous controversy in decision theory (the 'Allais controversy'), during which a methodology of normative falsification was first articulated and used to assess experimental results. In the third chapter, the methodological advantages of controlled experimentation are illustrated and discussed with examples taken from the experiments on the so-called 'preference reversal' phenomenon. Laboratory testing allows to establish with a high degree of certainty that certain phenomena lie behind the experimental data, by means of independent testing, elimination of alternative hypotheses, and the use of different instruments of observation. The fourth chapter is devoted to a conceptual analysis of the problem of 'parallelism'. This is the problem of inferring from the occurrence of a phenomenon in the laboratory, to its (possible) instantiation also in non-laboratory environments. Experimental economists have discussed parallelism at length, and their views are presented and criticised. Eventually, it is argued that parallelism is a factual matter and as such can only be established on empirical grounds. The fifth chapter provides an example of how one can argue for parallelism, focusing on the case of experimentation on the 'winner's curse' phenomenon. The role of experiments as 'mediators' between theoretical models and their target domain of application is illustrated, and the structure of parallelism arguments analysed in detail. Finally, in the last chapter, economic experiments are compared to simulations, in order to highlight their specific characteristics.

Acknowledgments:

The greatest debt goes to my supervisors, John Worrall and Philippe Mongin. Not only they have read all the preliminary drafts of this thesis, criticised them, and advised me how to improve them; they have taught me how to think philosophically in the first place. I should also thank all my other teachers at the London School of Economics, and in particular Dan Hausman and Nancy Cartwright, who have read extended parts of the manuscript and have strongly influenced my views on economic science. Various chapters have been improved thanks to Paul Anand, Roger Backhouse, Marco Del Seta, Donald Gillies, Brendan Larvor, Peter Lipton, Mary Morgan, Matteo Motterlini, Andrea Salanti, Bob Sugden, Sang Wook Yi, and a few anonymous referees. Members of the research students' seminar, the 'Modelling' and the 'Measurement in Physics and Economics' group at the LSE, and the Experimental Economics seminar at the University of East Anglia have also provided me with comments, suggestions, and a number of challenging critiques. I also thank audiences at the 'Economie et philosophie' seminar in Paris 1997, the University of Trento workshop on Economic Experiments 1998, the Model-Based Reasoning conference held in Pavia 1998, and the Italian Society for Logic and Philosophy of Science 1999, where some of the arguments have been tested. Thanks to all the researchers and staff at THEMA, University of Cergy-Pontoise, for their hospitality in the Spring terms of 1997 and 1998. While writing this thesis, I have been supported financially by a Marie Curie grant (TMR) from the European Community, a British Academy scholarship, and logistically by the Centre for Philosophy of Natural and Social Science at the LSE. The CPNSS has been a fantastic environment, for the invaluable intellectual stimuli it has provided, and for the people I met there. In particular, thanks Cynthia, Dorotha, George, Julian, Klementina, Kate, Makiko, and Sang. A special mention goes to Marco, Matteo and Sebastiano for their friendship, and for having diverted my thoughts from philosophy to football and other important matters during lunch breaks. Finally, I must thank Francesca and Francesca, who have made me happier.

Contents:

List of tables and figures:

Chapter 1

Introduction
Philosophy, experiments, and economics

'We must next deal with its style of presentation, and so cover both *what* is to be said and *how* it is to be said.'
(Plato, *The Republic*, Part 3, 392c)

1.1. The laboratory turn in philosophy and in economics

To quote famous economists on the irrelevance or even the impossibility of controlled testing is one economic experimenters' favourite games. One may start with John Stuart Mill: "there is a property common to almost all the moral sciences, and by which they are distinguished from many of the physical; that is, that it is seldom in our power to make experiments in them". "Our belief [in economic generalisations]" writes Lionel Robbins "does not rest upon the results of controlled experiments". According to Milton Friedman, "we can seldom test particular predictions in the social sciences by experiments explicitly designed to eliminate what are judged to be the most important disturbing influences"; and Richard Lipsey claims that "economics must be a non-laboratory science", given that "it is rarely, if ever, possible to conduct controlled experiments with the economy". Samuelson and Nordhaus say that "economists [...] cannot perform the controlled experiments of chemists or biologists because they cannot easily control other important factors", and even in the *Encyclopedia Britannica* one reads that "there is no laboratory in which economists can test their hypotheses".[1]

Nowadays, claims such as these have become rather obsolete. Economists *do* perform controlled experiments and *do* test their theories in the laboratory. Since the early sixties economics has been going through a 'laboratory revolution' like those that have irreversibly shaped other sciences in the past. Experimental economics is today an established sub-field of economics, experimental papers are published constantly in the best journals,[2] experimentalists hold chairs in the most prestigious departments, and some of them are even forecasted as possible Nobel laureates.

The laboratory revolution, however, is still 'in progress'. It is hard to say how deeply it has affected economic science, or whether it is just a temporary fashion that will leave no permanent mark on the discipline. Even sciences which have endorsed experimental practice as a basic methodology of enquiry, have done so to different degrees and use experimentation in different ways. In medicine, for example, laboratory experiments play an ancillary role in reaching conclusions that go *beyond* what happens in the laboratory. Tests on guinea-pigs and mice constitute an important aid while developing a new drug; but only the drug's capacity to defeat an epidemic in a population of human beings constitutes a valid test of its success.[3] Non-laboratory techniques of data-gathering survive in these sciences and are likely to live on forever. In contrast, a science like physics has been shaped by the laboratory revolution so deeply that it now has little in common with the Aristotelian science that it used to be (i.e. a science mainly concerned with understanding naturally occurring phenomena by means of unaided observation). Today - as students of science have noticed – most physics is devoted to the study of phenomena, such as cold fusion or lasing, that rarely if ever occur spontaneously in nature but are produced at will in the lab.

The laboratory has become a pervasive and characteristic feature of contemporary science, and philosophers and historians devote more and more time and effort to its study. A veritable 'laboratory revolution' in science studies

---

[1]Cf. Mill (1836, p. 124) and Robbins (1932, p. 74). The Friedman and Lipsey quotes are taken from Starmer (1999), Samuelson and Nordhaus' from Friedman and Sunder (1994, p. 1), the last one from Davis and Holt (1993, p. 4 n. 2)..

[2]According to Plott (1991), about a hundred experimental articles were published each year in the early nineties.

[3]On the 'laboratory revolution' in medicine, see the studies in Cunningham and Williams (eds. 1992).

has indeed occurred more or less simultaneously with the 'laboratory revolution' in economics. Philosophers like Ian Hacking and Nancy Cartwright, historians such as Peter Galison and David Gooding, sociologists like Bruno Latour and Harry Collins, have tried to save experimental practice from an earlier general 'neglect' (Franklin, 1986).

Both these revolutions, in economics and in philosophy, will loom large in this dissertation. The dissertation is *about* the laboratory turn in economics, and will make use of a number of conceptual tools developed during the laboratory turn in philosophy of science in order to understand it. My main interest being epistemological in character, I shall focus especially on the methodological and philosophical issues arising from the use of experimental methods to gain economic knowledge. Questions such as 'What kind of knowledge can experiments provide?', 'What does it mean to perform a good experiment?', 'What are the limits and advantages of experimentation?', will be the primary subject of interest throughout the following chapters. This of course will require some patience in order to understand the problems economic experimentalists address and the techniques they devise to solve them.

1.2. Experimental economics in a nutshell

The origins of experimental economics are difficult to trace back in time. Alvin Roth (1995, p. 4) points to Bernoulli's Saint Petersburg's experiments (1738)[4] as the earliest ancestors of today's experiments, but the question has little more than anecdotal interest. The truth is that experimental economics is almost entirely a post- Second World War product, and thus a fairly young discipline. For this reason, perhaps, we still lack a serious historiography of the subject.[5] What we do have is a set of personal recollections by the main protagonists of the 'laboratory turn', which have now become part of a 'standard narrative' encapsulated in many textbooks, handbooks and survey papers.[6] Customarily, three main lines of research are distinguished at the origins of experimental economics (cf. e.g. Davis and Holt, 1993, ch. 1; Roth, 1995). The distinction

---

[4] More about this 'experiment' in Ch. 2.

[5] To the best of my knowledge, the only historian who has written on experimental economics so far is Robert Leonard - see his (1994) article on the study of bargaining behaviour in economics and psychology.

[6] See Smith (1991; 1992), Davis and Holt (1993), Friedman and Sunder (1994), Kagel and Roth (eds. 1995), Hargreaves Heap and Varoufakis (1995). Smith (1992) and Friedman and Sunder (1994, ch. 9) provide to my view the best narratives in terms of attention to the institutional background and individual histories. Roth's introduction to Kagel and Roth (eds. 1995) is better in terms of number of references.

provides a very useful taxonomy of the kind of work done today (and this should perhaps make us a little sceptical regarding its historical accuracy).

The first line is that of *market experiments* (or *industrial organisation*). E.H. Chamberlin in the late forties began to experiment with his students at Harvard mainly for didactic purposes, in order to show them that even simple 'classroom markets' do not equilibrate as required by orthodox equilibrium theory (Chamberlin, 1948). Vernon Smith, a Harvard graduate student, was puzzled by some features of Chamberlin's experiments, and started to devise other, methodologically more sophisticated ones, when he moved to Purdue in the early sixties. His 1962 paper - "An Experimental Study of Competitive Market Behavior", which is now considered a classic - pointed to an innovative way of studying markets, and started a whole new literature on the subject. Smith is the main experimentalist working on markets and market institutions and one of the '*gurus*' of experimental economics in general.[7]

At the same time, a number of economists, psychologists and mathematicians were beginning to devise experimental tests of *game theory*. The list is very long: John Nash, Lloyd Shapley, John Milnor, Jacob Marschak, Sidney Siegel, Lawrence Fouraker, Martin Shubik and many others belong to it. They worked mainly in Princeton, Stanford, and the RAND Corporation in Santa Monica; others gathered around Reinhard Selten in West Germany. An interdisciplinary conference held in Santa Monica in 1952 was the first occasion for the American group to meet and stimulated further activity. Since then, game theory and experimental economics have become deeply intertwined and 'experimental games' constitute a substantial portion of contemporary game theory research.[8]

The third line of research concerns *individual behaviour* and *decision making*. Economists had been interested in the experimental determination of indifference curves since the early thirties. The real boost to experimental decision theory came however in the fifties, following von Neumann and Morgenstern's (1944) formalisation of an empirically testable model of rational decision under risk. At a conference held in Paris in 1952, Maurice Allais presented his experiments which - he claimed - refuted the von Neumann-Morgenstern model

[7] Smith's own historical account can be found in the introduction to Smith (1991), and in Smith (1981; 1992).

[8] For some history, see in particular Smith (1992) and Friedman and Sunder (1994). Mirowski (forthcoming, ch. 6) describes the context in which game theory (and experimental game theory) was born and flourished during and immediately after the war, with particular attention to the connections with military research.

at the descriptive as well as at the normative level. The history of this 'refutation' and of the subsequent debate will be told in more detail in chapter two (see also the references therein). After the Paris showdown, research in this area somehow stagnated until the late seventies, to finally become a most active branch of economics in the last twenty years.

The work done in each of the above areas of experimental economics has one feature in common: the attempt to gain economic knowledge by observing rather simple, repeatable processes in closed, 'controlled' circumstances. Friedman and Sunder (1994) distinguish experimental economics from other kinds of empirical testing precisely on the basis of the kind of data it uses. *Experimental economics uses experimental data deliberately created for scientific purposes in controlled conditions (the laboratory).* 'Happenstance' (as opposed to experimental) data are products of *uncontrolled* processes. 'Field' (as opposed to laboratory) data are gathered in 'natural' environments.

The notion of 'control' is central in the experimental sciences. At a very abstract level, experimental methods of enquiry consist in the variation of some (allegedly causally efficient) factor keeping all the other circumstances constant, so as to observe the effect of that factor acting alone on the system under study. By iterating such a procedure, the influence of all the causal factors and their contribution to the system's behaviour can in principle be studied, and different hypotheses tested. The laboratory allows one to carry on such an investigation in privileged conditions, in which 'background circumstances' can be kept constant, 'disturbing' factors 'shielded', and the 'main causes' triggered at will.[9] More concretely, to achieve control of some experimental system is a matter of skill and a heavily context-dependent question. Different techniques are applied from case to case, and it is difficult if not impossible to discuss them without some specific examples in mind. Some textbooks (see Friedman and Sunder, 1994, in particular) include guidelines about how to prepare, run, interpret an economic experiment, and even how to get the results published - at a level of detail that cannot be pursued in this introduction. In the course of this dissertation, however, I shall try to address concrete problems of experimental control and how they may be solved in practice.

---

[9] John Stuart Mill's (1843) four methods of 'agreement', 'difference', 'residues' and 'concomitant variation' constitute the first philosophical attempt to articulate the logic of experimental and quasi-experimental science.

What do experimental economists try to achieve? Most methodological papers include a taxonomy of the goals that may be pursued in the laboratory. Alvin Roth (1986; 1988; 1995) proposes three categories to classify economic experiments: 'Speaking to theorists', 'Searching for facts', and 'Whispering in the ears of princes'.

i. Speaking to theorists

A great deal of experimental economics has to do with testing and modifying theories which are already formalised and specified to a good degree. Most of the work on individual decision-making, for example, is of this sort. An axiomatised theory (the Bayesian model) was tested and some anomalies discovered; other models encompassing the old theory and the anomalous evidence were developed and tested on their own, and so on. Similarly, a lot of game-theoretic research follows this approach, ever since the early tests of behaviour in 'prisoner dilemma' situations. Experimentalists often argue that the laboratory enables to test theories in circumstances in which all their (explicit) assumptions are satisfied. The theories are, so to speak, given 'the best possible shot'. This methodology is supposed to have many advantages, compared to traditional tests 'in the field'. I shall come back to this issue in chapter four, where claims of this sort will be discussed and criticised.

ii. Searching for facts

Some economic theories are admittedly incomplete. This may be due to technical difficulties in solving very complicated models in an analytical way, or to the necessity of idealising from some aspects of the real world which cannot be easily represented. Many game theoretic models, for example, fall in the former category: some games have no equilibrium, others have many and we do not know which one is the correct prediction. An obvious solution to problems of this kind is to get some real human beings, let them play the game, and see what will happen. Experimental work on institutions, in contrast, falls in the second category. Neoclassical economic theory does not tell a very realistic story about the way equilibria are attained. Normally, a story of 'Walrasian *tatonnement*' is told, but no Walrasian auctioneer is at work in real markets. One of Vernon Smith's earliest achievements when he turned to the laboratory was to show that institutions matter and influence the outcome allocation more than previously assumed. In this case, experimental economics has filled a gap in economic theory providing data upon which, some day perhaps, a more complete account of the functioning of markets will be based.

iii. Whispering in the ears of princes

Other experiments are devised explicitly to support policy decisions. These experiments typically regard the effect of some institutional change, and are performed when no existing theory can illuminate the problem. Often some informal hypotheses - put forward by lobbies, industries, experts, etc. - are tested (as in type-i experiments). For example, whether a certain market institution leads to more efficient allocation than another. Alternatively, the experiments are used in order to solve problems that cannot be understood theoretically (as in type-ii experiments): can, for instance, a given auction system achieve certain policy goals, like promoting firms owned by minorities and women, and at the same time maximise the amount of revenue raised by the auctioneer?[10]

Smith (1982) uses also the 'Boundary Experiments' category. In this case the goal is to test the boundary condition under which the theory fails, and thus establish its degree of generality (robustness), or the possibility of a domain extension. Friedman and Sunder (1994, pp. 7-8), to conclude, isolate five major purposes of experiments in economics:

1. Influence a specific (policy) decision
2. Discover regularities where theory does not help
3. Test robustness (boundary experiments)
4. Choose among competing theories
5. Study the functioning of an institution before implementing it in the field.

The list mostly overlaps with the taxonomies found in other articles and textbooks. There is quite a general consensus among practitioners about the *scope* of experimental economics, but how are these goals achieved in practice? This is the central subject of this dissertation, which is mainly concerned with questions of *method*. I shall try to focus not only on general philosophical topics - such as the nature of economic knowledge, and the role of experiments in its production -, but also on the details of experimentation.

1.3. Philosophies of experiment

[10] An example of the use of experiments to solve problems of this kind can be found in Plott (1997). I have discussed the methodological sugnificance of Plott's work and its contribution to the design and implementation of the Federal Communication Commission auction of Personal Communication Systems in a separate paper (Guala, unpublished).

Philosophical rationalisations of experimental practice and of the role it plays in science are as old as experimental science itself. Galileo, Newton, Descartes, all had their own views about the proper use of experiments, views that were partly endorsed and partly revised by later scientists, historians and philosophers of science. Rather surprisingly, however, *detailed* accounts of experimental practice were until very recently almost non existent in science studies.[11]

This is not so any more: the so-called 'Philosophy of Experiment' (Hacking, 1989a) or 'New Experimentalism' (Ackermann, 1989) is one of the liveliest trends of research in recent philosophy of science. Fairly general presentations and more or less complete surveys of this growing literature can be found in Ackermann (1989), Hacking (1989), Pickering (1992), Knorr-Cetina (1992), Mayo (1996, ch. 2), Franklin (1998), and Morrison (1998a).[12] In the present and the following sections I shall sketch the main features of these new philosophical studies of experiments. The approach adopted varies from author to author, and it would be impossible to provide an exhaustive picture of this field of research in a few pages. There are however some fundamental tenets and general trends that characterise most philosophy of experiment, and I shall be content to focus on these. I shall start by illustrating the background - i.e. the 'traditional' philosophy of science that was dominant until approximately fifteen years ago - against which the philosophy of experiment represents an innovative turn. Then, I shall proceed to discuss the new literature. A very rough distinction can be drawn between those 'students of experiment' who endorse some form of scepticism or epistemic agnosticism about science and those who pay a tribute to its achievements.[13] The former correspond more or less to the 'sociologists of science', and - not the least for chronological reasons - it will be convenient to start with them (sections 1.5-1.7), before moving to other approaches (section 1.8).

1.4. The background: traditional philosophy of science

As noticed by Ian Hacking, in order to have an argument, and a fruitful one, two philosophers must share some common ground. Philosophy of science in the

---

[11] With a few notable exceptions: cf. e.g. Bernard (1865), Fleck (1935/1979).

[12] There are also a few volumes of collective papers, the most relevant being Achinstein and Hannaway (eds. 1985), Gooding, Pinch and Schaffer (eds. 1989), Le Grand (ed. 1990), Pickering (ed. 1992) and Buchwald (ed. 1995). Several journals have devoted special issues to this topic: see *Science in Context* (1988), *Isis* (1988), as well as the symposia in *PSA 1988*, *PSA 1990*, *PSA 1996*, and *Journal of Philosophy* (1989).

[13] Cf. Hacking (1989) for such a distinction.

English speaking world was dominated until quite recently by a tradition founded on a rather limited set of general claims. Hacking has pointed to a number of assumptions shared on the one hand by Carnap, Hempel, Nagel, Braithwaite and the other proponents of the 'received view'; and on the other by Karl Popper, one of their most vigorous critics. The 'ideal typical' traditional philosopher of science subscribed to some or all of the following principles: (a) scientific theories are mainly assessed in the light of empirical evidence; (b) science (physics at least) is in general progressing; (c) scientific language is precise; (d) science is in some sense unified; (e) the context of discovery and the context of justification are distinct; (f) the aim of philosophers is to isolate the rules for the rational justification of theories.[14]

These claims can be reformulated in a normative way: theories *should* be assessed mainly in the light of empirial evidence, scientific language *should* be precise, ... etc. At the core of the traditional approach lies the programme of isolating the features that characterise 'Science' and differentiate it from other activities such as, say, poetry or fishing. Once these features have been identified and formalised, a set of rules can be put forward that will tell us how 'good science' *ought* to be done.[15]

Attention to the history and practice of science is therefore entrenched in the traditional view. *Real* 'good' science provides in principle the 'empirical basis' against which theories of science should be tested.[16] In the late sixties, in particular, a sharp 'historicist turn' took place in philosophy of science, thanks mainly to Kuhn's *Structure of Scientific Revolutions* (1962/1970), and Lakatos' (1970) methodology of scientific research programmes. Philosophers turned to the study of the diachronic, dynamic aspects of science, and the new literature on experiments can be seen partly as an epiphenomenon of that renewed interest in the history of science.

The studies of experiment constituted, however, in the first place a new kind of *historiography* of science, aimed at illuminating aspects of the scientific enterprise that had been previously neglected. To begin with, a tendency to focus on bigger and bigger units of historical analysis - from theories to paradigms,

---

[14] I have slightly modified Hacking's (1983, pp. 5-6) list.

[15] For Carnap (1950) and the positivists this is the problem of *explication*, for Popper (1934/1959) it is the problem of *demarcation*; the two have slightly different characteristics which is not worth pursuing here.

[16] For a sophisticated presentation of the relationship between philosophy and history of science, see Lakatos (1971). Cf. also what is said below in sections 1.10-1.12.

research programmes or entire disciplines - was characteristic of the late traditional view, whereas the new studies of experiment opposed to it a programmatic attention to the *small* and *local* details of scientific activity.

Secondly, philosophers of science since at least the logical empiricists tended to focus on the *theoretical* and *linguistic* aspects of science. A crucial role, for instance, was assigned to 'observation statements', that were supposed to provide the low level empirical ground upon which the (theoretical) structure of scientific knowledge could be built. Rather than engaging in a study of how phenomena are *produced* in the laboratory, traditional philosophers were mainly concerned with how they are *described* by scientists, and with the logical relation between such descriptions and high level theoretical statements. The standard 'mechanisms' of science, according to the traditional view, are the deductive-nomological (D-N) and the hypothetico-deductive (H-D) models:

Theory
Initial Conditions

∴ Observation statement

Where is the experiment in such a scheme? It is behind the observation statement (O), it is so to speak what 'produces' O. Pierre Duhem (1905), and many others after him, had pointed out that in order to obtain the 'right' O we have to make sure that the apparatus works well; since we implicitly formulate a hypothesis about the functioning of the instruments, traditionalist philosophers introduced among the premisses of the D-N and the H-D models a concern for experiments in the form of auxiliary hypotheses: that the theories of the instruments are correct, that everything works well, etc. It may sound a little exaggerated, but this is more or less all you can find in traditionalist philosophy of science about experiments, and this reduces to talking about *theories* of instrumentation, *hypotheses* about the correct performance of an experiment, etc.

Even philosophers and historians who challenged the standard view seemed to share the 'linguistic' prejudice of the traditional approach. Norwood Russell Hanson (1958), Thomas Kuhn (1962/1970) and Paul K. Feyerabend (1975), for example, stressed the importance of the 'background' assumptions that influence scientific observation, but in so doing highlighted once more the theoretical aspects of science and ignored the material, practical ones. The first detailed study of laboratory work, to be sure, shared this 'literary bias' with the authors cited above. Yet, Latour and Woolgar's *Laboratory Life* (1979/1986) marked a turning point in recent studies of science.

1.5. The sociology of science

The earliest studies of experimental practice were done by sociologists, and it would be impossible to understand those studies without introducing the basic ideas informing the new 'sociology of scientific knowledge' (SSK) of the seventies. The so-called 'Strong Programme' in the sociology of science, developed in Edinburgh by David Bloor (1976) and Barry Barnes (1974), can be taken as the 'manifesto' of the SSK. Their main claim was that a *principle of symmetry* should be applied to historical explanations of scientific development: sociological explanation does not apply only to by-products, external interferences on science, cases of fraud, pseudo-scientific disciplines, etc., but also to scientists' greatest achievements.[17]

The Strong Programme aimed at showing that - instead of methods of rational justification - *sociological factors*, in the last instance, *are what fix scientific beliefs*. In order to do so, it exploited three weaknesses of traditionalist projects of individuating rational rules of justification. In particular: the problem of underdetermination, the problem of theory-ladenness, and the lack of a general theory of confirmation.

I shall not address these issues in detail here. Some of them will be discussed in the following chapters – for example, theory ladenness and underdetermination in chapter three. For the purposes of this chapter, it is important to notice that the Edinburgh School took as units of analysis wide time-periods and whole cultural phenomena. Sociologists 'of the laboratory', in contrast, focused on very local events and limited periods of time. Theirs has been labelled the 'micro' approach to SSK, as opposed to Bloor and Barnes' 'macro' approach. From this point of view, it marks a sharper discontinuity with traditional philosophy of science. Micro SSK began to study 'laboratory life'.

1.6. The sociology of the laboratory

[17] Such a claim was directed in particular against the 'rationally reconstructed' history of science advocated by Lakatos (1971) and his pupils (see, e.g., the case studies in Howson, ed. 1976).

In *Laboratory Life*, Bruno Latour and Steve Woolgar (1979/1986) looked at a small community of scientists as an anthropologist would study a tribe, and described the events in the laboratory as they resulted from the interaction of the agents, their negotiations, the social conventions that regulate their behaviour. One outcome of scientists' work is the 'discovery' of scientific facts, which are said to be '*constructed*': they are the result of a number of operations performed by actors constrained by social rules, conventions, and institutions. Without such rules, institutions, conventions, etc., the 'facts' would not be there, would not be (accepted as) facts, or would be interpreted in a different way.

Latour and Woolgar studied the discovery-construction of a thyrotropine releasing hormone (TRH) by a group of scientists later to be assigned the Nobel Prize in medicine. In order to do so, they held an *agnostic position* on scientific knowledge, arguing that an anthropologist should be neutral towards the form of life she is studying and restrain from applying categories that constitute the very object under investigation. Latour and Woolgar, therefore, did not pronounce on the status of statements produced in the laboratory, and tried not to use epistemic categories that played a role in the settlement of the controversy over TRH.

The authors of *Laboratory Life* pointed in particular to the macro-approach's inability to account for the overwhelming importance of the 'technical' in the fine structure of laboratory work (1979/1986, pp. 25-27). The advertised attention to *practical* aspects, *non-verbalised* actions, and the importance of the *material* and *technical* apparatus in the laboratory, however, took a rather strange form in Latour and Woolgar's work, who took 'literary inscriptions' as the basic unit of analysis for the sociologist of science. The production of scientific papers was taken as the main goal of scientific activity, and the widespread use of inscriptions (notes, data, drafts, articles) a key feature of scientific work. Latour and Woolgar distinguished between five types of statements corresponding to different degrees of 'facticity'. The goal of a scientist is to produce so-called '*type 4*' statements, i.e. uncontroversial, explicit claims corresponding to new, uncontroversial scientific facts (1979/1986, pp. 76-81). The shifting of a statement from one type to another is a matter of *negotiation* between the agents working in the laboratory, as well as between them and the outside social environment. The construction of scientific facts is 'social' because the decisive factors for the assignment of a 'type-4 facticity' to a statement are (allegedly) social.

As it appears especially from the second edition of *Laboratory Life* (1986), however, *anything* is 'social' from the anthropologist's perspective (1979/1986,

p. 281). ‘Construction’ is “the slow, practical craftwork by which inscriptions are superimposed and accounts backed up or dismissed” (1979/1986, p. 236). The game played by the scientific tribe being “directed [...] toward [...] operations on statements” (p. 237), every step in the construction can be seen as ‘craftwork on statements’ rather than on real entities. Gaston Bachelard’s (1953) original idea of *‘phénomenotechnique’*, or that phenomena cannot exist without the apparatus used in their production, was therefore severely weakened by Latour and Woolgar. The technical, material aspects of laboratory work, apparatus and technology, were treated as ‘inscription devices’ (1979/1986, p. 51), statement-producing machines. The principal units of analysis being literary inscriptions, the approach adopted in *Laboratory Life* seemed unable to highlight the distinctive traits of science as opposed to other statement-producing activities: knowledge is constructed and science is *just one* kind of construction among many others (1979/1986, p. 31).

Laboratory Life was nonetheless a pioneering text where many themes crucial to the later literature were first addressed, albeit not always discussed in depth. Latour and Woolgar, for instance, challenged the traditional habit of studying ‘ready-made’ products of scientific activity, usually theoretical, often formalised accounts as they are published in scientific journals, whose “traces of production are made extremely difficult to detect” (Latour and Woolgar, 1979/1986, p. 176). By rejecting the traditional distinction between context of justification and context of discovery, sociologists have started a most fruitful enquiry into the preparation and implementation of experiments, with their complex apparatus and teams of scientists that make it work. Harry Collins is the sociologist who has spent most time studying and illustrating the difficulties of laboratory work, by applying the concept of ‘tacit knowledge’.

1.7. Tacit knowledge and underdetermination

The expression ‘tacit knowledge’ is due to Michael Polanyi (1967), and denotes the set of allegedly inarticulable skills, unverbalised abilities that make the practice of ‘good’ science impossible for those who have not received an appropriate training. Tacit knowledge may be a challenge to those interested in scientific methodology. Some students of science (and notably traditionalist philosophers of science) believe that it is a realm of phenomena *in principle* analysable and articulable; others, like Polanyi, believe that a thorough explication of scientific practice is impossible - and this *irreducibly* tacit dimension would undermine the explication and demarcation projects at their roots. If it is impossible to isolate and define precisely a set of criteria to distinguish 'good' from 'bad' science, science is merely what 'good' scientists do. Criteria of appraisal would be private, and scientific activity could not be assessed in the light of intersubjective standards.

Harry Collins has built on Polanyi to argue for what he has called 'the experimenter's regress'.[18] Discussing in particular the case of the discovery of gravitational radiation, Collins (1985, ch. 4) argues that when there does not exist a clear benchmark to arbitrate between 'good' and 'bad' experiments (such as a preliminary knowledge of the properties of the sought-for phenomenon), scientists are typically caught in a circle of validation between experiments and results: "when the normal criterion - successful outcome - is not available, scientists disagree about which experiments are competently done", and "when there is disagreement about what counts as a competently performed experiment, the ensuing debate is coextensive with the debate about what the proper outcome of the experiment is. The closure of a debate about competence is the 'discovery' or 'non-discovery' of a new phenomenon" (1985, p. 89).

Collins's case studies are alleged to show how a high degree of arbitrariness follows from such a circularity: scientists might have trusted an apparatus that had been rejected very early in the controversy, or the other way around. Finally, it is possible to argue from arbitrariness to the social construction of science. The order that we believe characteristic of natural phenomena is merely a historical contingency, the result of negotiation and a chain of arbitrary decisions.

It is worth discussing the argument for social construction in *Changing Order* because it is a standard one we find in most micro-sociological analyses of scientific activity. The scientist is initially facing some task, e.g. the construction of a complicated apparatus such as a laser. His work typically proceeds on the basis of a set of presuppositions, tacit conventions, background knowledge, acquired skills, and goals. The most characteristic feature of laboratory work is that nothing goes smoothly. Experimenters proceed by trial and error, and at each stage a number of possible strategies are open to them. A typical list of options includes: to give up and just accept that the task is in fact impossible (for instance, because those who claim to have successfully performed the experiment earlier lied - recall the cold fusion case); to decide that the apparatus is responsible for the failure,

[18] The notion of experimenter's regress has sparked quite a big controversy; cf. Franklin and Howson (1984), Collins (1985), Franklin and Howson (1988), Radder (1992), Franklin (1994) and Collins (1994).

modify the instrument; to decide that there must be something wrong in the theory about how the effect obtains, modify the theory; to decide that the problem is due to a lack of the experimental skills, go and get more information or training; to decide that the problem lies in the analysis of data, use other techniques; simply to try again hoping that the next time it will work, etc. The history of an experiment looks like a chain of decisions taken in highly uncertain circumstances.

The next step in the sociologists' argument is a formulation of the well-known (logical) principle of underdetermination of theory by data: empirical evidence *alone* cannot settle controversies nor establish changes in conceptual order. The question then arises quite naturally: *why* do scientists do what they do? (SSK is a scientific discipline, looking for explanations.) Old and new epistemologists provide a number of answers to such a question, trying to answer the underdetermination challenge in a logical or methodological fashion. Some of these solutions will be illustrated in chapter three, and their relevance to understand the development of a real controversy in experimental economics discussed in detail. SSK students share the view that no answer of this sort is correct, and thus point to a different direction.

Step three is to claim that scientific debates, *in practice*, are closed by convention. It is through conventions that the social comes into play. Collins' depicts the situation by means of a metaphor of 'networks' of concepts and practices (1985, p. 131).[19] The idea is that of a holistic structure where everything is related to everything else. Changes in the network cause 'reverberations' that potentially affect any element in the structure. If reverberations are not welcome by some elements in the net, they will be resisted. An analysis in terms of 'costs' and 'benefits' is usually adopted at this stage in order to explain why some changes (discoveries) are accepted while others are not: the scientist can choose whether to challenge more and more links in the network, and (here is the step from conventionalism to sociologism) his choice is guided by 'social convenience'.[20]

Accepted scientific facts, then, result from a sort of *equilibrium* between all the actors and the forces at work in the situation at hands. When the system equilibrates, the product of scientific activity becomes a 'fact'. Andrew Pickering (1995b) talks of a process of 'tuning' leading to a 'robust fit' between the

---

[19] Andrew Pickering, similarly, speaks of 'matrices' (1984) and more recently of a 'mangle' (1995).
[20] Cf. e.g. Collins (1985), p. 138; Latour and Woolgar (1979), chs. 5-6.

components. The concept of 'equilibrium' fits well Latour and Woolgar's (1979/1986) description of scientific activity as a sort of bargaining.

A scientific fact, then, is what comes out of such a process of adjustment in the network. It is - to use Kuhn's terminology - what 'meshes' with the scientific paradigm. The SSK paradigm is typically a *theoretical* entity - a set of preconceptions, presuppositions, commitments, interests, metaphysical beliefs, and goals. Sociologists of science, thus, have often been read as claiming that nature is causally inefficient in the production of scientific knowledge; that the order we admire in nature is *superimposed on* nature and is merely a historical contingency, the result of negotiation; that the order we create is arbitrary and artificial.[21]

1.8. The epistemology of experiment

The sociological literature on experiments has typically been received by philosophers in either one of two opposite ways. On the one hand, many readers have been impressed by the general philosophical conclusions contained in these works: epistemic relativism and constructivism have been used to support postmodern, feminist, and critical positions of various sorts. On the other, philosophers concerned with the traditional projects of demarcation and explication have generally seen the sociology of scientific knowledge as a threat, and fought it with all the means at their disposal. But of course these are not the only options. A small number of philosophers, for example, have seen in the new studies of experiment an invaluable source of 'empirical data' on science.[22] The new sociological literature provided in fact students of science with accounts of scientific practice carried on at an unprecedented level of detail. It was possible, disregarding the overall conclusions contained in SSK studies, to build on these accounts in order to try to supersede both the traditional views, and the new philosophy of the constructivists.

Ian Hacking's *Representing and Intervening* (1983) was probably the first book in the philosophy of science to adopt such a strategy. One can find in Hacking's work as well as in subsequent studies an attention for traditional issues, such as the problem of realism of the unobservables, the problem of

[21] Latour and Woolgar suggest this point in the chapters dedicated to 'reification' and to the discussion of order vs. disorder (cfr. 1979/1986, chs. 1 and 6).

[22] Some sociologists seem to look at such an approach in a rather sympathetic fahion - see Collins (1991b) and Latour and Woolgar (1979/1986, p. 280).

explanation, and the problem of how to produce reliable scientific knowledge. A common theme in such literature is the attempt to overcome the traditional view and at the same time to avoid the pitfalls of constructivism, *by downplaying the role of theory in science*. This is done in a number of ways, and I shall try to provide a brief overview in the next few pages. (Some of these ideas will be discussed further and sometimes put into practice in later chapters, so what I am saying here should be taken as an appetiser for what is to come.)

(a) Science is more than theory testing

Popper is often quoted for his claim that "theory dominates experimental work from its initial planning up to the finishing touches in the laboratory" (1934/1959, p. 107). Popper noticed that even low level observational claims involve general terms such as 'lever', 'screen', 'microscope', etc. which presuppose some theories about the dispositional behaviour of levers, microscopes, screens, etc. Hacking has pointed out that such a view is true only in its most trivial sense. Any scientific claim involves theoretical presuppositions *of some sort*; but no particular insight is gained by calling all such presuppositions 'theoretical' without distinction. On the contrary, by so doing one will open the door to the holism of paradigms, with its alleged relativistic consequences.

A first useful distinction to be drawn is between 'high' theory (or 'Theory' with capital 'T') and 'low level' experimental assumptions and hypotheses. The main difference lies in the first being *explanatory* of the phenomena observed in the laboratory, whereas the latter constitute the 'background' against which phenomena are individuated. Experimental assumptions typically involve hypotheses about the functioning of the apparatus, about the absence of interferences, about the correct application of statistical techniques of data analysis, and so on. Theory (capital 'T') will in contrast provide a causal account of the phenomena observed, possibly in rigorous and mathematical form.

Once such a distinction has been introduced, it is clear that not all experiments are tests of some 'Theory'. Many involve mere calibration of the instruments, refinement of a yet unexplained effect, or exploration of a new domain of phenomena.[23] The distinction is important because experiments performed with different goals in mind are appraised in the light of different standards, and produce knowledge of a different kind.

[23]For some examples see e.g. Franklin (1986; 1990), Galison (1987) and Hacking (1983); on exploratory experiments, cf. also Steinle (1997).

(b) Experimental and theoretical knowledge are often independent

A number of philosophers have pointed out that the main goal of experimenters is *to establish phenomena*, rather than to test theory. They have invoked an independent epistemology of experiment, a set of rules and standards applied when the main concern is to make sure that an experiment has been *performed* correctly - rather than to check whether a certain phenomenon has been *explained* correctly. Allan Franklin (1986; 1990) has stressed that the *elimination of experimental errors* should be the main subject of such an epistemology.[24] Franklin (1990, p. 104; 1998) has compiled a list of techniques customarily applied in order to detect errors and make sure that what is observed is a real property of the natural world rather than an illusion produced by our techniques of investigation. (I shall present and discuss Franklin's list in chapter three below.)

The historian Peter Galison (1987) has similarly argued on the basis of a number of examples taken from the physics of small particles that most experimentation can be characterised as the attempt to isolate an effect from the 'noise' that surrounds it. This is done, roughly speaking, in two different ways: either by 'calculating' or by 'shielding'. In the former case, the contribution of the causal factor of interest to the observed data is spotted by eliminating the noise on paper, by means of calculations. In the latter case, in contrast, the effect is 'carved out' from the background by means of materially constraining the environment so as to prevent the disturbing factors from contaminating the data.

The fact that background elimination can often be performed even when the phenomenon is imperfectly understood accounts for the fact that experimental results have, to use Hacking's slogan, 'a life of their own'. Moreover, from a historical point of view, changes in Theory do not necessarily coincide with changes in experimental knowledge. The trajectory of the laboratory sciences is often independent of the developments of high level theory. Such a thesis has been used by different authors (see Hacking, 1983; 1992; Ackermann, 1985; Galison 1987; 1997; Buchwald, 1993) in order to question the monolithic character of Kuhnian and sociological paradigms. The objectivity and progessiveness of science, in fact, can be defended only if a continuous, rather stable empirical basis underlies changes in Theory. Philosophers of experiment

[24] See also Galison (1987), Hon (1989) and Mayo (1996).

believe that experimental knowledge can provide such a firm bedrock, surviving drastic revolutions in theoretical explanation. As opposed to the traditionalists, though, the bedrock is not constituted by observational statements (only) but by a set of instruments and techniques for the elicitation of data, together with some ‘home truths’ (low level causal claims) about their functionings, and eventually some historically robust ‘phenomenological laws’.

(c) Empirical data and high theory are not deductively connected

Another way to vindicate the independence of experimental from theoretical knowledge is to challenge the traditional view about the connection between Theory and observation. According to the standard accounts of explanation and theory testing (the D-N and the H-D models), observational statements are *deduced* from theory, initial conditions and auxiliary assumptions. More recent accounts tend in contrast to stress the role of heterogeneous models in the explanation of data, building in particular on Patrick Suppes’ (1962) semantic approach. Suppes argued that several layers of models connect high Theory to observation. These usually include models of the theory, models of the phenomena, models of the data, and models of the experiment, none of which is logically entailed by any other model.[25]

The role of models is a central theme also in Nancy Cartwright’s *How the Laws of Physics Lie* (1983) and in her later work. Cartwright argues that the construction of a good model of the phenomena involves the skilful application of a number of techniques of data-reduction, approximation, and idealisation.[26] Theory is *one* among many tools applied in constructing empirical models. Moreover, it is in models of the phenomena that the empirical content lies. Low-level models are true of the phenomena (if anything is), whereas abstract laws are strictly speaking false (and hence ‘lie’ about the real world) due to their incompleteness, abstractness, or idealised character. Bogen and Woodward (1988)[27] have synthesised some of these ideas in their distinction between *data* and *phenomena*. Given that such a taxonomy will be used extensively in chapter three, I do not want to spend too much time on it here. Suffice it to say that the reason for a distinction lies in phenomena being neatly and rigorously explicable,

---

[25] For an illustration of Suppes’ view with some examples, cf. Mayo (1996, ch. 2).

[26]On the procedure of ‘preparing a description’ upon which to theorise, see also Lynch (1990) and Gooding (1990).

[27] See also Woodward (1989) for further arguments and examples.

whereas data are mostly unpredictable and deductively inexplicable from high Theory.

(d) Science is intervening as well as representing

Finally, experiments are quite independent of theory (singular) in the sense that they depend on theor*ies* (plural) of different sorts. In contrast with the logical positivist accent on the unity of science, a renewed interest for its disunified character has followed, once again, Suppes' (1981) proposal. Some authors have turned the disunity of science thesis into the exploration of a metaphysical alternative (Dupré, 1993). Others have used the *fact* that science is disunited in order to argue for the objectivity of scientific knowledge.

Having recourse to different theories in order to interpret empirical data, in fact, does not necessarily involve any kind of vicious circularity - as long as the theories used to back up an experimental result are independent from those used to explain the result itself.[28] The disunity of science, then, is 'healthy' as long as it provides us with a set of tools to independently support scientific results.[29]

The argument from independence plays an important role also in Hacking's argument for the realism of unobservable entities. In *Representing and Intervening*, Hacking points to the obvious fact that our belief in science derives at least as much (if not more) from our being able to use science to manipulate the world than from our trusting the representation of the world provided by it. Applying such an intuition to the problem of realism, Hacking (1983, ch. 11) argues that belief in the reality of what is observed through microscopes comes (1) from the fact that different microscopes (polarising, electron, fluorescence, light microscopes, etc.) present the same visual structures; and (2) that when we manipulate the object under study the visual data change in a way that is consistent with the existence of an object with certain properties.

Both are typical 'no miracle' arguments: it would be an incredible coincidence if instruments working according to completely different principles produced the same illusory effects; and it would be a sort of cosmic conspiracy if every time we intervened on an object, the expected effect took place without our basic causal beliefs being at least approximately correct. In a subsequent chapter (ch. 16),

[28] Cf. Kosso (1989) for a rigorous discussion of this argument.
[29] See also Hacking (1996) and the whole volume by Galison and Stump (eds. 1996).

Hacking exploits the argument from intervention in a rather different way. Unobservable entities can not only be manipulated, but also used as tools in order to study *other* aspects of nature. In order to hunt free quarks, for example, small particle physicists make use of tiny magnetised balls of a material called niobium. The charge is varied by spraying the niobium ball with positrons and electrons, so that different measurements of its level of transition from positive to negative charge become possible. According to Hacking, "If you can spray them, then they are real" (1983, p. 23). We believe in electrons and positrons not (only) because we believe in theories of the electron, but because we think we can build reliable instruments that, we think, use electrons to perform a number of tasks. "Engineering, not theorising is the best proof of scientific realism about entities" (1983, p. 274).[30]

(e) The 'material' plays a crucial role in science

Rudolf Carnap (1934) once introduced a famous distinction between two alternative ways of speaking about a given subject. To speak in the '*material* mode' is to put forward claims directly about extra-linguistic objects. To speak in the '*formal* mode', in contrast, is to put forward claims about linguistic entities only. Philosophers of science have for a long time (more or less consciously) followed Carnap's advice that philosophy should deal with the logic of the language of science, leaving extra-linguistic objects to science itself. Thus, they have studied science focusing on what scientists *say* about reality (and about the instruments used to investigate it). Philosophers of experiment have urged that we turn to reality itself and speak about it 'in the material mode'. Such a view is held in common by a number of scholars of different views, and is an obvious way of further downplaying the role of theory and possibly avoiding the paradoxes of social reductionism.

A standard strategy to challenge the Strong Programme, in fact, is to try a sort of *reductio ad absurdum* by pointing out that reality cannot be merely reduced

[30] Hacking's argument raises two worries. First of all, it restricts the set of entities one is allowed to believe in to just those one can play with in the laboratory. Hacking (1989b) has chosen to bite the bullet and indeed reject realism about non laboratory entities such as galaxies and black holes. Secondly, the argument from intervention is based on the assumption that there exist low level causal relationships that are known by the experimenter but are independent from high theory. Morrison (1990) has questioned such an assumption, showing by means of concrete cases – including those presented by Hacking – that 'low level' causal relationships are most often embedded in, and dependent on, highly sophisticated theoretical frameworks. In order to be interpreted correctly, then, interventions seem to require a lot of reliable Theory in the background.

to *whatever* is consistent with the prejudices held by the community. Such a form of idealism (social prejudices, conventions, ideologies, etc. are after all *ideas*) seems inconsistent not only with the history of science (which is full of examples of theories that have clashed against 'hard facts' and have thus been abandoned no matter what their proponents believed or desired); above all, it goes against common sense: reality is not the slave of our desires and cannot be shaped at will.

Such an intuition is of course central in older accounts such as Popper's falsificationism (according to which reality plays the pivotal role of saying 'no' to human attempts to capture it by means of bold theoretical hypotheses), but had been severely weakened in the light of the well known problems of falsificationism and of the holistic theses of Duhem and Quine. A renewed interest in the 'material' aspects of science has recently stimulated an interesting dialogue among constructivists and epistemologists of experiment. David Gooding, in his *Experiment and the Making of Meaning* (1990), has argued on the basis of a number of case studies in the history of electromagnetism that the goal of laboratory work is primarily to achieve a stable interaction between the experimenter(s) and the real world. Focusing on the concept of human *agency* in the material world, Gooding claims that "the world and our representations are made to converge by the activity of the observers" (1990, p. 86).

Gooding shows by means of examples that the 'creation' of a phenomenon involves a manipulative activity on reality as well as, simultaneously, the construction of a series of models of the phenomenon. When these two activities converge by mutual adjustment and a stabilisation is achieved, the phenomenon is established, the practices are routinised, and experimental knowledge diffused. Gooding's examples fit well the ideas of Ian Hacking and Nancy Cartwright, according to which 'preparation of materials', 'creation of phenomena', and 'preparation of descriptions' account for a great deal of everyday scientific activity.[31]

Andrew Pickering has recently (1995a; 1995b) begun to talk about science in what he calls a 'materialist idiom'. Pickering's up-to-date version of the network is called the 'mangle'. In a number of case studies, he tries to show that in the history of an experiment no decision is dictated by any specific element of the

[31] According to Cartwright (1993; 1995), indeed, models play a crucial role in science because theoretical laws are true principally of models and of those real world situations that are artificially constructed so as to resemble the models. Very similar views can be found in Bhaskar (1975) and Latour (1984; 1986).

mangle of practices, beliefs, interpretive accounts, resistance of the material world, rigidity of the instruments, etc. *Contingency* seems to reign: it 'just happened' that things went as they did; things 'could have developed otherwise'.

Hacking has endorsed some aspects of Pickering's account and spoken of a new version of 'Duhem's thesis', in a 'materialist mood'. When something does not match our expectations, we not only modify our representations of the world - we can, and do, also modify the world. Using Duhem's example: we not only reformulate a theory of the instrument, we modify the instrument (Hacking, 1992, §11). What is a scientific fact, then? For the traditionalists, knowledge is true justified belief (or rationally justified belief, if you are a fallibilist). For the SSK, it is what meshes with a framework of social habits, practices and beliefs. For Gooding, Pickering and Hacking, it is what results from the adjustment of theories, practices, and the material world. Scientific facts are established by mutual interaction of a number of components, including reality, materiality, the technical (cf. Hacking, 1992, p. 58).

Hacking notices however that some theories appear historically robust because only they can make sense of the 'rigidity' of the apparatus and of the 'resistance' of nature. "The common experience of the laboratory sciences is that there are all too few degrees of freedom. (...) It is extraordinalily difficult to make one coherent account, and it is perhaps beyond our powers to make several" (Hacking, 1992, p. 55). Like Hacking, Peter Galison (1987; 1997) has challenged sociologists's conclusions by means of the notion of multiple *constraints*. Galison - in opposition in particular to Pickering's (1984) accent on contingency - stresses the rigidity of the network the scientist is working within. The debate goes on, and it is not my intention to contribute to it here.[32]

1.9. Experimenters' philosophy

An obvious question may be asked at this stage: why should all these controversial views be of relevance to the present topic? To begin with, they have been developed in the context of philosophical debates concerning the natural sciences, and it is not clear why they should matter to economics. Secondly, economists surely have developed their own approach to experimentation, and it would be wise to start a philosophical study of experimental economics by taking a look at what they say about their own work.

[32] See in particular Galison (1987; 1995; 1997), Pickering (1995a; 1995b), Hacking (1999).

Concerning the first objection (economics is not physics) let me postpone a reply until the next section, where I shall illustrate and defend my views on 'methodological pluralism'. About the second one, I can easily concede that it is most sensible to look at how economic experimentalists see their own work. The rise of experimental economics took place against the background of a generally hostile profession, and therefore experimentalists felt the need to justify their approach to a degree that is nowadays quite unusual among economists and perhaps even scientists in general. (Few economic *theorists*, for instance, would waste their time justifying the utility of their highly mathematised models, sometimes of very remote practical applicability.) Experimental economists, as a consequence of this situation, are on average more self conscious of the limits and possibilities, pros and cons of their activity than most of their colleagues.

Still, we lack a thorough philosophical discussion of experimental economics. Most of the literature on the methodology of experimental economics comes from introductory sections of scientific papers, preliminary methodological chapters in textbooks, handbooks,[33] and a few printed debates between experimentalists and/or their critics.[34] A third category of publications includes papers written for purposes of popularisation, dictionaries, inaugural lectures and other special occasions.[35] Methodology has been a constant preoccupation to just a few authors - Vernon Smith, Alvin Roth and Charles Plott, to name just the most authoritative ones.

Professional philosophers have generally speaking neglected experimental economics.[36] This is the opposite of the trend observed among economists, who are more and more attracted to the experimental approach. Why has it been so? Probably because it is difficult to construct a suitable pair of 'philosophical' spectacles through which experimental economics can be analysed in an interesting way. Traditional philosophy of science, as we have seen in section

---

[33]Hey (1991); Davis and Holt (1993); Friedman and Sunder (1994); Kagel and Roth (eds. 1995).

[34]See e.g. Harrison (1989) and the following discussion in the *American Economic Review*, 82, (1992), as well as Kagel and Battalio (1980) and Cross (1980)

[35]Cf. Smith (1987a; 1987b; 1989), or Plott (1991).

[36]The second volume of the journal *Economics and Philosophy* featured a paper written by Alvin Roth (1986), thus pointing to an area of potential interest for philosophers of science and methodologists of economics - but nothing followed. Other isolated attempts to start a dialogue were made by Wilde (1981) and more recently Starmer (1999). With the notable exceptions of two studies on the preference reversals anomaly (Hausman, 1989; 1992; and Tammi, 1997 – more on this in chapter three), however, philosophers have not considered experimental economics a subject worth of research.

1.4., takes experimental practice merely as an 'evidence-producing' machine, functional to the aim of verifying/falsifying theories. Nothing intrinsically interesting is supposed to happen in the laboratory.[37]

Given that we lack a general philosophy or methodology of experimental economics, we are left with the scattered ideas that can be found in the texts cited above. When very general claims are put forward concerning the role of experiments in economics, they are usually taken rather uncritically from traditional philosophy of science.[38] The most valuable contributions, in my view, concern rather *specific* methodological puzzles experimentalists have faced in the course of their activity – especially problems that have been traditionally neglected by philosophers of science.

The main (and the *best*, I think) example of this kind is provided by the debate on 'parallelism'. I shall talk extensively about it in chapters four and five, because I take it to be a rare example of a genuine methodological problem that (a) has been discussed at length by experimental economists, (b) has been generally neglected by philosophers of science, and (c) can teach us something about the use of experimentation in other disciplines than economics. In chapter four, in particular, I shall examine critically what experimentalists have said about parallelism, and try to state the problem in a revised, conceptually more rigorous fashion.

I shall also address practitioners' views at length in chapter two. There, again, I shall be concerned with a set of philosophical issues that are quite peculiar to economics. I shall in fact try to articulate decision theorists' 'methodology of normative testing', a set of rules used to appraise normative theories of rational choice in the light of experimental evidence. Clearly, such a problem can only arise in a hybrid - half descriptive and half normative - science like economics. In order to isolate these rules, I shall examine the methodological controversy that, in the early fifties, set the French economist Maurice Allais against the defenders of the standard view.

---

[37] Not surprisingly, no particularly valuable insight came out of the interaction between experimental economics and traditional philosophy of science: cf. Smith, McCabe and Rassenti (1991). Harry Collins' reaction to this paper will be discussed at length in chapter seven.

[38] Cf. for instance Friedman and Sunder (1994, p. 2).

I shall not take experimentalists' views uncritically but will try to assess them and sometimes improve on them. But, one might ask, from which standpoint? How can one criticise methodological propositions (claims, moreover, put forward by people who are supposed to be experts in the subject in question)? These worries revolve around the problem of how to do philosophy of science in general, and methodology of economics in particular.

1.10. Methodological monism and pluralism

The philosophy of economics has always oscillated between two extreme approaches: methodological monism and methodological pluralism. *Methodological monism* is the doctrine according to which the production of knowledge follows a unique set of rules and principles common to *all* disciplines. Economics and physics, human and natural, formalised and non-formalised, experimental and non-experimental, micro and macro, fundamental and non-fundamental sciences – according to such a standpoint – fall under the same methodological and philosophical 'umbrella'. *Methodological pluralism*, in contrast, asserts that different methods of enquiry are appropriate to different disciplines, because of their peculiarities in terms of subject, level of analysis, content, aims, level of complexity, etc.

This methodological dichotomy has taken different forms in different periods and depending on the authors who have addressed it. In its most famous version, it has concerned the natural and the human sciences, and has been debated by great social scientists and philosophers such as Weber, Marx, von Mises, Dilthey, Robbins, Knight, von Wright, von Hayek, Gadamer, Quine, Davidson and many others.

Rather than taking the issue of monism vs. pluralism at face value – something that would require a PhD dissertation on its own – let me briefly defend the position I shall adopt in the following chapters by contrasting it with some other approaches that have been popular in the recent past. During the last hundred years or so, at a surface level, the methodological gap between economics and the natural sciences (physics in particular) has been considerably reduced. Like physics, economics has become highly formalised and mathematical;[39] it has extensively adopted statistical techniques of data analysis;[40] it has made use of

[39] On the influence of mathematical physics on economic thought, see in particular Mirowski (1989).
[40] On the birth and development of econometrics, cf. Morgan (1990).

Monte Carlo and computer simulations;[41] economists are keen to be seen as respectable scientists and have their own Nobel Prize; they have adopted a strong empiricist rhetoric; and, moreover, have repeatedly tried to apply philosophical doctrines originally developed in the context of the natural sciences to their own problems. Logical positivism, falsificationism, Lakatosianism, Kuhnianism, Feyerabendism, postmodernism have all been popular at different times among economists.

The direct transfer of philosophical doctrines from the natural sciences to economics is motivated principally by the apparent power of the natural sciences. Physics is the paradigmatic example of a science that 'works', whatever that means (of course philosophers argue about that as well). If any science has achieved the capacity of efficiently manipulating the external world, it is physics; if humans have achieved any genuine knowledge of nature, it must lie in physics; if any discipline has progressed, it must be physics. If we are looking for a set of standards characterising 'good' science, so the argument goes, we should start taking a look at the best science available. The methods applied in a 'strong' discipline like physics, in fact, are most likely to constitute a good recipe for doing 'good' science in general.

It is not surprising, then, that almost every time economics (ar another 'weaker' discipline) has entered a state of crisis, some economists have turned to the methodology of the most developed sciences looking for inspiration. In the thirties, in the middle of the Great Depression and at the dawn of Keynesianism, economists looked at neopositivism and (in one case) at falsificationism;[42] in the seventies, facing the decline of Keynesian macro-models, they turned to Kuhn and Lakatos in order to make sense of what was happening and possibly find a way forward.[43]

These applications have been fruitful to the extent that they have provided new frameworks for the interpretation of the more or less recent history of the field, and have stimulated economists to ask new questions about the meaning of their own work. Thanks to the philosophy of science, economists have tackled important issues such as: Do our models (theories, laws, axioms, etc.) have

---

[41] The historical and methodological literature on simulations in science (and in economics in particular) is scarce. Cf. the works cited in chapter six below.

[42] Cf. Robbins (1932) and Hutchison (1938).

[43] For some early applications, cf. Ward (1972) (on Kuhn and economics) and the essays collected in Latsis (ed. 1976) (on Lakatos and economics).

empirical content or are they mere formal structures, tautologies, interpretive frameworks? Does economics aim at discovering the causes of social phenomena? Have economists been successful in predicting events in the past? Is it important to anticipate novel facts in order for a science to be progressive? More fundamentally, what does it mean to make progress in economics?, and so on. The debates that have revolved around these questions have surely increased our understanding of economic science, and by comparing economics to stronger disciplines (or to the more or less idealised picture of such disciplines provided by philosophers of science) we have clearly gained some insights about its limits and potentialities.[44]

1.11. The unbearable lightness of economic methodology

Yet, such methodological monism seems unable to satisfy the needs of the philosopher eager to learn from economics (as opposed to the economist or historian of economics interested in understanding his or her own discipline). The problem can be exemplified by means of the following *dilemma*. Suppose we compare some methodological doctrine originally developed in the context of the natural sciences (let us call it M_n) with economists' practice. We shall find ourselves in either one of the following two cases:

(a) if economic practice satisfies at least approximately or for the most part M_n's methodological precepts, then economics gets good marks while the normative value of M_n (grounded on the unquestioned power of the natural sciences) is little enhanced;

(b) if, in contrast, economists' practice does *not* satisfy M_n's standards, then economics scores low in the game of science, whereas the value of M_n is left equally intact.

The dilemma, in other words, stems from the fact that economics seems to be too weak a science either to constitute the 'empirical basis' for an autonomous methodology (that is why contemporary monists turn to natural science methodology in the first place) or to 'falsify' any methodological doctrine developed in the context of a stronger science. If economics does not work very

[44] A number of authors, for example, have assessed positively the attempts to apply Lakatos' methodology of scientific research programmes to economics and the history of economic thought. See Hands (1990, p. 79), Backhouse (1994, pp. 186-188), and Salanti (1994, p. 31).

well (or at least not as well as other disciplines such as physics), then it is idle for the philosopher of science to look at economics in search of philosophical inspiration. To look at physics straight away is a much better strategy. Economics is too weak to bear the weight of methodological construction.

Such a consideration has rather devastating effects, especially in the light of those theories of methodology (or 'meta-methodologies') that assign a crucial role to scientific practice in the assessment and improvement of philosophical theories of science. Lakatos' (1971) 'virtuous' circle of justification between scientific practice, history of science, and philosophy of science – to take a well-known example – does not even start spinning in the case of economics. If in economics (as recognised by someone sympathetic with openly normative philosophy of economics) "it is difficult to agree on a list of undisputed scientific achievements", and "it cannot be taken for granted that what is commonly regarded as best practice is directed towards discovering the truth" (Backhouse, 1994, p. 184), then we are left with the natural sciences as the only 'normative basis' for our methodology. As a matter of fact, Lakatos himself argued that "a good methodology - 'distilled' from the mature sciences - may play an important role for immature and, indeed, dubious disciplines" (1971, p. 137, n. 4).[45]

1.12. How to (dis-)solve the dilemma

In the light of such reflections, what role is left for economic methodology? In this section I intend to show how the dilemma above can be, if not solved, perhaps *dis*solved. The arguments presented in the previous section rely on the assumption that *economics is weak*. It must be stressed again that such an assumption has origins that can be easily traced back in time: periods of intense methodological and philosophical debate coincide almost invariably with periods of deep crisis in economics. Sometimes the crisis is opened by the rise of a new paradigm replacing an old and established one; but most often the crisis is 'external' to the discipline, arising from economists' incapacity to anticipate and deal with major economic events such as the Great Depression of the thirties or the persistent stagflation of the seventies.

[45] My position on Lakatos' methodology and on its role in appraising economic theories has been influenced and stimulated by many years of discussion with Matteo Motterlini. Matteo, however, would not necessarily agree with all my conclusions (see Motterlini, 1999).

In such situations, economists' inferiority complex towards other disciplines like physics, which appear to proceed from one predictive and technological success to another, becomes deeper than ever. But this feeling of inferiority may be more a contingent effect of such critical circumstances than the result of a rational and systematic comparison between economics and the natural sciences. My proposal to get out of the dilemma, then, is to reject the background assumption that economics is a 'weaker' science than (and therefore has to learn methodology from) physics.

In order to question the assumption of inferiority, I shall begin by refraining from talking about economics *as a whole*. To speak in too general terms is the best way to fall back into the dilemma. John Sutton, in a forthcoming book, follows a similar strategy and focuses on a set of situations in which economic models seem to work well - not much worse than many physical theories that are customarily cited as exemplars of scientific excellence. In this dissertation I shall be even more cautious and concentrate on a rather small bit of economics – experimental economics – giving up the ambition of making general claims about the scientific status of economics in general.[46] I take this attitude to be consistent with the message of recent philosophy of science, a philosophy that urges us to abandon big units of analysis and to concentrate on singular cases with their local, specific, contingent peculiarities. By proceeding this way, of course, one faces the risk of missing the 'big picture' (with its big and very important questions) and getting lost in useless, detailed narratives of local events of minor importance. I think that the risk is worth taking, and bet that it will pay off in terms of a more realistic, fuller, more exciting picture of what contemporary science is and of what economists can (and cannot) do. As in several other cases, of course, the proof of the pudding is in the eating.

A third reason - beyond escaping the dilemma and achieving descriptive accuracy - to endorse localism is strictly philosophical in character. Traditional philosophy of science was engaged in the bold project of articulating a universally valid, content-free, purely analytical scientific Methodology (capital 'M'). It aimed, in other words, at a purely formal set of inferential rules analogous to those of deductive logic. Had such a formal system been found, the validity of a scientific inference could have been assessed independently of the situation, the

[46] Of course I cannot help but having some views about economics in general, and such views will surely be rather transparent to the reader who will be patient enough to get to the end of this dissertation. They should not, however, be taken as the fundamental message of the present work.

content (empirical truth) of the propositions involved, and any contingent facts of the matter.

Such a hope has now been abandoned, for reasons that were already *in nuce* in Pierre Duhem's writings, were clear to Otto Neurath and have more recently revived by Quine. The problems of induction and underdetermination have severely undermined the project, in a way that for reasons of space I can only illustrate by means of an example. The riddle of induction is often taught to philosophers by means of a short story widely known as 'Russell's chicken': "The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken" (Russell, 1912, p. 98). Russell's chicken had very good (inductive) reasons to believe that the farmer was carrying food on the fatal day, as he had always done. One of the morals that can be drawn from this story, and from the history of the inductivist programme, is that a general rule of inductive inference and no matter how much evidence will not suffice to produce reliable knowledge. Much more information of a substantive sort is needed to avoid the chicken's fate. Had the chicken known about the general context in which its actions were taking place, and of the background conditions underlying the regularity in question, he could have done better. The same moral can indeed be drawn from the Duhem-Quine problem and used against purely deductive methodologies such as Popper's falsificationism. Notice that problems of this sort undermine the idea of a universal and purely formal scientific Method, not the methodological enterprise itself. In our chicken world there *are* correct inferences to be made every single day; but evaluating which inference is correct will depend on many contingent facts, not least on how the world is. If we want to do normative methodology, we better study in an *empirical* way the *local, contingent, 'refined'* methods (small 'm') applied by scientists case by case, giving up the hope of a universal Method valid a-priori.[47]

Of course no one denies that there are circumstances in which economics does *not* work well; the point is, this should not push us to throw too much (say, the neoclassical paradigm, or even economic science as a whole) away.[48] Even the most advanced sciences like physics can predict with precision the behaviour of

[47] For an outline and defence of the argument in relation to the philosophy of the social sciences, see Kincaid (1996, ch. 2). Cartwright, Cat, Fleck and Uebel (1996) trace the origins of this approach to methodology back to Neurath's work.

[48] Sutton (forthcoming), for example, is very critical of general equilibrium modelling and the influence it has had on economic theory (e.g. on the new classical macroeconomics).

only a limited number of systems while struggling in vain with important phenomena that still require an explanation. According to an old analogy, indeed, the social phenomena economics is concerned with are more similar to natural phenomena like the weather than to the phenomena that have been successfully tamed by physicists. They are similar, of course, from the point of view of *complexity*.

The striking successes of physics are to a great extent due to the possibility of dealing with simple, closed, tightly controlled and 'artificial' systems. Were the capacity to predict storms the standard of appraisal for physics, physics would not look much better than economics. When one cannot tightly control and manipulate the initial conditions of the system under study (no matter whether the system is a physical or a social one) rather poor knowledge is going to be produced. If we look at the history of an old and famous problem such as the explanation of the perihelion of Mercury, for example, we see that it has evolved in several respects like many economic controversies, with periods of relative agreement followed by new harsh controversies that questioned the old results. The observation and explanation of most astronomical phenomena is indeed affected by problems of sparse and poor data, constant interferences, impossibility of collecting new evidence, difficult measurements, controversial theoretical assumptions, etc. to such an extent that every result is imbued with radical uncertainty and a consensus is rarely if ever reached by the community of experts.[49] Thus, the problem with economics - as with many branches of the natural sciences - may lie in the complexity of the subject matter rather than in some alleged methodological inferiority. Despite its simplicity and plausibility, such an argument has been surprisingly unpopular among philosophers of economics.[50] (It is more likely to be found in the mouths of some economist or econometrician apologising for his or her predictive failures.)

Sutton (forthcoming) shows by means of a number of examples how economics works reasonably well in simple situations that are structurally similar to those described in its models. Experimental economics provides us with a further test of the 'complexity' hypothesis: if situations in which economic models work could be created in the laboratory with a certain ease, the differences between physics and economics would appear less dramatic than it has often

---

[49] The history of the controversy on the perihelion of Mercury is told in a non-technical way by Will (1988, ch. 5). For a recent study of the 'radical uncertainty' of macroeconomic modelling and the 'interpretative flexibility' of economic data, cf. Evans (1997).
[50] But see Hausman (1992) for a notable exception.

been assumed. I think that this conclusion can be legitimately drawn from the examples that I shall examine in the chapters below. Indeed, I believe that some form methodological pluralism can be defended but is 'orthogonal' to the traditional distinctions between economics and physics, '*Naturwissenschaften*' and '*Geisteswissenschaften*': it is much more interesting to focus on distinctions like laboratory vs. non-laboratory, mathematical vs. non-mathematical, statistical vs. non-statistical, etc. Different techniques are applied in different areas of a same discipline, depending on the problem situation, and providing results with varying degrees of reliability.

1.13. The strategy of this dissertation

My strategy, to sum up, will be the following. I shall look at a number of cases of experimental 'good' practice - at cases in which economics 'works'[51] - and use them as paradigm examples of how experimental economics can (and should) be done. One could reply that the choice of examples betrays an overenthusiastic attitude towards my subject. But the goal here is not to advertise experimental economics: for any positive example chosen, one can surely find some 'bad' experiment that has been devised and published by a respected economist. I shall leave this exercise to someone else. Whenever I shall think appropriate, I shall also highlight the methodological defects of the experiments discussed, point out that they do not necessarily or entirely support the claims the authors put forward in their papers, and sometimes even try to suggest how they might be improved.[52]

A radical empiricist approach to philosophy of science will be followed, and the arguments grounded as much as possible on cases of real scientific practice. From this point of view, I shall be faithful to the way of philosophising that has characterised the best traditional philosophy of science as well as much of the new philosophy of experiments. I shall try as much as possible to look at what economists *do*, as opposed to what they *say* they do, keeping in mind that "most scientists tend to understand little more *about* science than fish about hydrodynamics" (Lakatos, 1970, p. 62 n. 2).

---

[51] At this stage I can only leave an analysis of what 'to work' means to later chapters. As I shall show especially in chapter three, 'to work' can be identified with the elimination of alternative explanations of the data; in Guala (unpublished), with technological success; but there surely are several other measures of success in economics and in science in general (see Backhouse, 1997, for an attempt to pin them down).

[52] See in particular chapters four and five.

The approach will be mostly normative, the principal aim of scientific methodology being to understand how genuine scientific knowledge is produced and perhaps to help scientists to clarify (if not improve) their methods of enquiry. To isolate, make explicit, explicate, articulate, and possibly formalise methods that are often only implicit and unconsciously applied by scientists – that is the philosopher's task.[53]

I shall be cautious about the extent to which economists' methods of experimental enquiry can be generalised outside their own domain. To reiterate once more, a taste for 'localism' is one of the legacies of recent philosophy of science. In chapter three, however, I shall argue that the strategies adopted by experimental economists in order to tell real phenomena from artefacts do not differ from the methods used by the natural scientists, and that all such strategies fall under two general arguments that are well known to philosophers of science. In chapters four and five, I shall also suggest that the problem of parallelism, that has been a constant preoccupation to experimental economists, should be of concern to other scientists too.

I shall use in a number of cases ideas put forward by old and new philosophers of science. From the new literature on experiments, in particular, I shall take the suggestion to focus on the details of experimenting, the idea that theoretical aspects are less important for science than philosophers have tended to assume, and the taste for 'localism' and 'pluralism'. I shall *not*, in contrast, deal with some of the typical controversies that one finds in that literature: the debates between constructivists and realists, between those who see the history of science as characterised by radical discontinuities vs. those who see it as a cumulative process,[54] between those who stress tacit knowledge vs. those who see experimenters as applying rational and formalisable methods of enquiry.[55]

---

[53]Some philosophers believe all normative, meta-theoretical enterprises to be radically misguided and philosophically hopeless (cf. Rorty, 1979). Such views have become quite popular among economists (see McCloskey, 1985), but a serious attempt to criticise them would take me too far away from the main subject of this thesis. Similarly, I shall not try here to defend in detail the naturalistic approach to epistemology that I adopt. On these issues, at any rate, I happen to share many of the views put forward by Dan Hausman in his (1992, pp. 263-269 in particular).

[54] See for example Galison (1987; 1995; 1997), Pickering (1995a; 1995b), Hacking (1989; 1992; 1997), Stump (1996).

[55] Cf. e.g. Franklin and Howson (1984; 1988), Collins (1985; 1994), Franklin (1994).

Economic experimenters (just like experimenters in general) do several different things.[56] I have tried to take such variety into account as much as possible when choosing the examples and case studies. In chapter two I examine experiments devised in order to test and develop normative theories of behaviour; in chapter three, experiments aimed at establishing the producibility a phenomenon in the laboratory; in chapter four, experiments to validate the existence of a phenomenon outside the laboratory; in chapter five, experiments to reproduce in the laboratory a phenomenon that has allegedly occurred outside of it.

The theories at stake vary from case to case too: models of individual decision under risk will be central in chapters two to four, whereas game theoretic models of auctions will be prominent in chapter five. It is also worth making clear from the start that some interesting aspects of experimental economics have surely been neglected. In this thesis one will not find very much, for instance, concerning purely game theoretic experiments,[57] border-line experiments between economics, cognitive science and psychology,[58] or experiments on animal behaviour.[59] Experimental economics has become such a big business that to give an exhaustive picture of what is going on in the field (if possible at all) would require much more space than allowed in a PhD dissertation.

1.14. The content of this dissertation

In the *second chapter* I shall try to tackle a rather general question, which is of crucial importance for the purposes of this dissertation. Economics has traditionally had an ambiguous status. To some, it was a science totally grounded on *a-prioristic* foundations; others have interpreted it as a normative science of rational behaviour. According to these views, of course, very little room is left for an experimental branch of economics. I shall show how, in the early fifties, the French economist Maurice Allais (one of the fathers of experimental economics) got involved in a heated controversy with the defenders of Bayesian decision theory, about the status of their model and of the experimental counterexamples which apparently ‘falsified’ it. The ‘Neo-Bernoullians’ - as Allais named his opponents - began to defend their theory from a *normative* perspective. Instead

---

[56] See also section 1.2 above.

[57] See Camerer (1997) for a discussion and survey.

[58] See for example Egidi (1995) and the exploratory experiments done at the Experimental Lab of the University of Trento in general.

[59] Cf. Kagel, Battalio and Green (1995).

of giving up altogether and simply challenging the received view on a descriptive basis, Allais insisted that his counterexamples could prove the inadequacy of Bayesian decision theory also from a normative respect. During the controversy, a true 'methodology of normative testing' was developed, and the history of decision theory shows that Allais was right in pursuing the battle on both fronts. The moral I shall draw from this case is that normative commitments are crucial in economics. Models of individual decision making have to satisfy certain precise rationality requirements before they become accepted by the community of scientists. Still, this leaves a lot of room for experiments, *both* on the normative and on the descriptive side. I shall illustrate the point by means of examples from the history of decision theory, at the same time introducing some basic technical concepts that will be used in the rest of the dissertation.

In the *third chapter* I shall focus on a famous empirical anomaly, known as the 'preference reversal' phenomenon. The latter is a counterexample to all theories of individual behaviour under risk that satisfy a very basic principle of consistency - the transitivity axiom. The phenomenon was first 'produced' in the laboratory by a group of psychologists, and only later addressed by economists. A number of attempts to 'explain the phenomenon away' as an artefact of the experiment proved to be unsuccessful, and now preference reversals are generally accepted by economists as an established experimental fact. I shall use this case to illustrate how different techniques can be applied in order to test different explanations of the experimental data. These techniques do not differ substantially from those used in other laboratory sciences, and I shall also devote some time to illustrating the relationship between the problem of artefacts and more traditional philosophical puzzles such as the 'theory-ladenness' and the 'Duhem-Quine' problems. Experiments allow one to establish a phenomenon up to a degree of certainty that is rarely achieved in other branches of the social sciences. Eventually, agreement becomes almost unescapable even in a notoriously litigious discipline like economics. These, I shall argue, are the 'pros' of experimentation.

I shall turn to the 'cons' in *chapter four*. Economic experiments in fact have problematic features that do not seem to arise in other fields. In particular, established experimental results are not automatically generalisable to non-experimental domains. Given that economics is in great part devoted to the study, explanation, and control of non-laboratory phenomena, this problem is a serious one. Experimentalists have called it the problem of 'parallelism': how can experimental results be generalised from the laboratory to real-world economies? I shall discuss the problem from an abstract point of view, and at the same time review some arguments for 'parallelism' that have been put forward by practitioners. Finally, I shall propose a new formulation of the problem and also a solution to it. I shall use the case of preference reversals again in order to illustrate how a rigorous analysis of the problem may provide guidance to experimentalists in devising and interpreting their experiments.

In *chapter five* I shall build on the previous discussion in order to sketch a general characterisation of the role of experiments in economics. Experiments, I shall argue, are just an intermediate step in the testing of an economic hypothesis: they act as 'mediators' between theoretical models and the target systems whose behaviour they aim at explaining. The argument is illustrated focusing on experiments on the 'winner's curse' phenomenon. In this case the tests were devised in order to check whether the occurrence of a certain phenomenon allegedly hidden behind 'field' data could be established in a tight fashion. After the 'winner's curse' was rather uncontroversially (re-)produced in the laboratory, experimentalists made use of field data in conjunction with experimental evidence in order to argue that the parallelism step could be taken with confidence. I shall also suggest that parallelism arguments are analogical in character.

The last chapter - *chapter six* - is devoted to some general comments and conclusions. I shall discuss a general issue concerning the nature of experiments in economics. What is the difference between a genuine experiment and a simulation? And in which category do economic experiments fall? Developing an insight due to Herbert Simon, I shall argue that a distinction between simulations and experiments can be drawn neatly from a conceptual point of view: a 'pure' experimental system is made of the same material as the target system of interest; a simulation in contrast makes use of different 'stuff' in order to reproduce properties similar, at a certain level of abstraction, to those of the target system. In practice, however, economic experiments are often hybrid objects sharing both experimental and simulating features. This is not necessarily a problem, because experiments and simulations are used in a very similar way.

Chapter 2

Testing Normative Theories of Rationality

The debate on expected utility theory

> 'And does a man remain at unity in himself in all these experiences? We saw that there could be conflict and contrary opinions about the same objects in the realm of vision; isn't there a similar conflict and internal struggle in the realm of action?'
>
> (Plato, *The Republic*, Part 10, 603c-d)

2.1. Introduction

The status of economics among the empirical sciences is a notoriously controversial issue. To begin with, economists' desire to be considered empirical scientists is a relatively recent phenomenon. Secondly, as I have briefly mentioned in the introduction, it is not clear whether empiricist commitments go much further than mere rhetorical statements.[60] Thirdly, even if economists were really for the most part empiricists, they would be so in a way that departs from most traditional empiricist philosophical accounts.[61]

Be that as it may, I am not concerned with economics *as a whole* here, but with a rather limited portion of it - namely, experimental economics. Experimentalists are indeed professed and practicing empiricists, who turned to the laboratory in order to acquire a better empirical basis for economic science. Yet, the success of experimental economics depends very much on the issue

---

[60] See Blaug (1980/1992) and McCloskey (1985), two authors who disagree about more or less everything but this particular issue.

[61] But see Hausman (1992) for a recent attempt to square economics with Millian empiricism.

above: were economists (in general) not empiricists, the work of the small laboratory community would have little if any impact on the discipline as a whole.

The strongest and most die-hard antiempiricist doctrine in economics has probably been so-called '*a priorism'*. According to a priorists, economic theories are assessed on grounds that are independent of (and 'prior to the collection' of) empirical evidence.[62] Such grounds may be of various sorts: old a priorists like Menger, von Mises and Lionel Robbins took the fundamental principles of economics as Aristotelian 'essences', Kantian 'categories', necessary conditions for the analysis of the subject matter. Modern a priorists in contrast tend to assess theories in the light of their mathematical elegance, or their consistency with a certain economic tradition (for example, in the light of their using equilibrium analysis rather than not).

According to a most influential tradition, indeed, economics is defined as 'the science of *rational behaviour*'. Vilfredo Pareto, for example, characterised pure economics as the study of "the many logical, repeated actions which men perform to procure the things that satisfy their tastes" (1906, ch. 3, §1). Such a view is by no means of mere historical interest: claims of a similar sort can be found in modern textbooks, and economics is often praised for its normative power - rather than for its purely descriptive capacity to explain how real economies work.[63] 'Normativism' is particularly strong in specific areas of economics that have grown and proliferated in close contact with experimental economics, such as game theory and decision theory. But if game and decision theorists are mainly concerned with rational decision making, isn't the experimental collection of evidence about *real* decision making an idle activity?

I shall approach this important question by examining a particular debate that started in the early fifties and continued with fluctuating intensity for about three decades. The debate was sparked by the French economist and future Nobel prize winner Maurice Allais, one of the founders of experimental economics. Allais, as I shall show in more detail shortly, claimed to have falsified by means of experimental data a model of rational decision making well known to every student of economics - the Expected Utility (EU) theory that had been formalised

[62] Notice that according to my definition Mill, Cairnes and Neville Keynes do not belong to the a priorist tradition because they believed in introspection as a method of theory validation. Blaug (1980/1992) calls them in fact 'empirical a priorists'.
[63] Cf. for instance Kreps (1990, p. 7).

a few years earlier by von Neumann, Morgenstern, and a number of other economists and mathematicians.

This alleged ‘falsification’ provoked a reaction from the supporters of the EU model, who tried to resist Allais’ conclusions by defending their theory from a *normative* standpoint. Allais could have simply shuddered and taken the theory as falsified from a descriptive point of view. Instead, he questioned the normative adequacy of EU theory as well, and started a controversy that ended up with the abandonment (in the decision theory community, at least) of the old model and its replacement with other ‘generalised’ theories of decision under risk. But what does it *mean* to ‘falsify’ a normative theory? The issue is not entirely clear even nowadays, and surely it was not clear at all when Allais and his opponents started their debate. Their controversy, thus, was as much about the value of their theories as about the methodology that should be applied in order to appraise them.

In the first part of the chapter (sections 2.2-2.3) I shall reconstruct this controversy in order to show how a ‘methodology of normative appraisal’ slowly emerged from the debate. Since experimental evidence plays a crucial role in such a methodology, I shall use this case to argue that even if disciplines like decision theory (or game theory, in the context of which the EU model was originally developed) were entirely concerned with normative questions, there would still be room for experiments in them. A more realistic picture of these areas of research, however, involves a *mixture* of normative and descriptive considerations. In the second part of the chapter (sections 2.4-2.5), thus, I shall try to show how certain basic normative principles must be built into every model aspiring to be accepted by the profession. Such principles act at the same time as a sort of ‘precondition’ for the acceptance of a theory, and as ‘heuristic guidelines’ for its construction. Again, this supports Allais’ intuition about the importance of fighting his battle at both the descriptive and the normative level.

One disclaimer: historical accounts of the development of decision theory are almost non-existent, and mine should not be taken as an attempt to fill this gap. The narrative I shall propose focuses on a very limited set of features of the story - the debate about normativism and the role of rationality principles in post- von Neumann and Morgenstern developments of decision theory. There are several aspects that are not addressed in this chapter but which surely have played a crucial role in directing research in this area: the difficult relationship between economists and psychologists, and between experts in decision theory, management scientists and general economists, or the influence of military-funded operation research during the war, just to name a few.

My reconstruction nevertheless highlights some aspects of the story that have been neglected by the existing accounts. The latter have customarily tried to separate the descriptive from the normative concerns of the protagonists, usually by focusing on the former and omitting the latter. Fishburn and Wakker (1995), for example, have produced an analysis of the different formulations of expected utility theory as they emerged in the pioneers' writings. Their story, by focusing on the formal aspects, abstracts completely from the dialectical contrapositions, both on descriptive and normative issues, that provided the impetus for theory-development. Mongin (1988a) puts forward a fuller picture by reconstructing the debate as a successful struggle with the problem of identifying the assumptions responsible for empirical refutations to the received theory. Decision theorists are portrayed as good empiricists applying a reasonable strategy in order to solve a methodological puzzle, the Duhem-Quine problem, ubiquitous in the natural and social sciences. The solution, according to Mongin, was dictated by the nature of the empirical falsifiers and by the formal structure of the falsified theory.

My version of the story aims at improving on formalist ones by concentrating on the problem-situation lying behind theoretical change. In particular I shall focus on the strict relation between the problem of the *interpretation* of the rationality principle and the problem of its *normative justification*. Are there many principles of rationality, or is there a unique legitimate candidate? And what is the criterion for their assessment? I shall not tackle these questions directly, but illustrate how they were discussed in the history of contemporary decision theory. I should make clear right from the start that the approach of this chapter will be mainly descriptive. Although the issues discussed below raise very important methodological questions (for example: is it legitimate to let normative considerations constantly influence descriptive ones?), I shall mainly focus on decision theorists' peculiar logic of discovery: the problem of normative justification will be shown to play a crucial role in guiding the development of alternative descriptive theories of human behaviour under risk. My account also aims to improve on 'descriptivist' reconstructions by showing that it is impossible to make sense of much of the debate without taking normative questions into account in the framework of analysis. It is also intended as a tentative contribution to the largely unexplored field of economic heuristics.

In the second section I shall put forward a 'potted' history of early expected utility theory, from Bernoulli to von Neumann and Morgenstern and the emergence of the first 'paradoxes'. Then, I shall analyse the defensive strategies adopted by orthodox theorists in order to explain away the counterexamples, and the debate between Allais and the 'American School' on the foundations of rational choice theory (section 2.3). In the fourth section I shall discuss an alternative model of decision, Machina's 'generalised expected utility analysis' and its normative status. Finally, it will be shown how weaker and weaker informal notions of rationality have been employed in order to justify more radical departures from the received model.

2.2. The rise of expected utility theory

According to the *Oxford English Dictionary*, a 'logical paradox' is "a statement or proposition which, despite [apparently] sound reasoning from an [apparently] acceptable premiss, leads to a conclusion that is against sense, logically unacceptable, or self-contradictory".[64] It is not, however, the sense in which the term has usually been used by decision theorists, as we shall see. Common denominators of paradoxes are self-contradiction, absurdity, or just the property of being contrary to "received opinion and belief" and "preconceived notions of what is reasonable or possible". According to official historiography, expected utility (from now on EU) theory was generated by a mathematician in order to solve a paradox of probability theory: the famous Saint Petersburg paradox, proposed in 1713 by Nicholas Bernoulli to the mathematician Montmort.[65] Imagine tossing a fair coin until a head appears; if the first head occurs at the *n*th toss you will win 2^n dollars: for any integer n you have a probability of 2^{-n} of winning 2^n dollars. The paradox arises from the consideration that if one follows the apparently 'reasonable' rule of maximising the expected monetary payoff of a lottery, he should be ready to pay *any* sum to enter such a lottery with a potentially infinite payoff. But, of course, no reasonable man would give all his money to participate in the Saint Petersburg game. It is a true logical paradox: an unreasonable conclusion is derived from rather plausible premisses. In 1738 Daniel Bernoulli suggested to replace the premiss with a more realistic one: agents aim at maximizing 'moral values' rather than sums of money, and since the 'moral value' of the monetary gain decreases as n increases, it is possible to find a finite value for the game. Such a solution, by creating EU theory, provoked also the

[64]For the same definition of 'paradox' with a number of examples, see also Sainsbury (1988).
[65]For a history of the Saint Petersbug paradox, see Jorland (1987).

first shift in the theory of rational decision under risk. Denote a typical lottery by $x = (x_1, p_1; ...; x_n, p_n)$, where the x_i refer to the outcomes and the p_i to the associated probability values. Bernoulli replaced the conjecture that in case of random lotteries agents act so as to maximise the mathematical expected monetary gain

$$(0)\ \ EG = \sum p_i x_i$$

with the conjecture that agents maximise the mathematical expected ‘moral value’

$$(1)\ \ EU = \sum p_i U(x_i).$$ [66]

Bernoulli's 'moral value' was later assimilated to the notion of 'utility', formally representable by a number assigned to an outcome, and intuitively corresponding to the introspective degree of 'pleasure' and/or 'pain' experienced by agents should that outcome occur.[67] Neoclassical economists (Jevons, Edgeworth, as well as - though not so explicitly - Marshall) conceived utility as a cardinal measure of happiness independent of the context (whether risky or not). Typically, the metaphysics underlying classical cardinalism takes utility to be a ‘psychophysical’ entity (pleasure/pain) and leaves no or little role to the preference concept. But for our purposes we may gloss over this fact and introduce a notion of cardinalism applicable both to the earlier writers and some modern economists: this notion postulates that utility differences represent differences in the intensity of preference among certain outcomes. Taking $(x, y) >> (z, w)$ to mean that the difference in preference between two objects of choice x and y is greater than the difference in preference between two other objects z and w, a cardinal utility scale must satisfy the condition

$$(1^*)\quad (x, y) >> (z, w) \Leftrightarrow u(x) - u(y) > u(z) - u(w).$$

In the light of his rationale for the adoption of (1) - namely, that moral value diminishes at the margin - Bernoulli arguably took utility as a cardinal measure of

---

[66] Bernoulli proposed a definition of utility as a logarithmic function of wealth. For an English translation, see Bernoulli (1738/1954).

[67] Modern utilitarians rank pains and pleasures on a same scale, in order to make possible an arithmetic calculation of the best solution available. Bentham, on the other hand, distinguished between different ‘circumstances’ (metrics) of pleasures and pains (see Mongin, 1995). Early utilitarians were *not* primarily concerned with problems of rational individual decision, but rather with problems of welfare economics, and thus the 'best' solution was defined as that which maximizes the happiness of all individuals in the society.

satisfaction in a riskless context.[68] However, as we shall see shortly, the expected utility formula can also hold for another utility representation, thus defining another notion of cardinality than the 'difference in preference' one.[69]

Cardinalism was to be challenged on the philosophical point that it is not at all clear how to compare (introspectively or otherwise) two differences in utility. Pareto showed as early as 1898[70] that in fact cardinality is not required in order to obtain the theorems of neoclassical economics, and just conjectured the existence of an ordinal utility function, which was later shown to be equivalent to assuming merely asymmetric, transitive, and continuous preferences. Pareto was mainly concerned with the fact that only indirect evidence could be presented in favour of the existence of a cardinal utility scale, which could not be directly and intersubjectively measurable. An ordinal 'index', by contrast, could be obtained by direct observation of economic choices. By dismissing cardinality, the 'ordinalist' revolution which followed Pareto[71] was to have significant consequences for EU theory, as explained shortly.

In (1926), Ramsey - in an initially almost completely ignored paper - sketched an 'empirical' basis for the axiomatisation of EU theory. The full axiomatisation, as well as the systematisation and diffusion of decision theory under risk and uncertainty,[72] came only in the forties with the work of von Neumann and Morgenstern (1944) (vNM from now on), who adopted the Bernoullian hypothesis of maximisation of mathematical expectation (1) above.[73] Ramsey had

---

[68]Cf. Bernoulli (1738/1954): "any increase in wealth, no matter how insignificant, will always result in an increase in utility which is inversely proportional to the quantity of goods already possessed" (p. 25). But the issue is not clear: according to Jorland (1987, p. 161), Bernoulli "comes close to define units of utility in terms of probability".

[69]In a purely mathematical sense, an index is cardinal if unique up to a positive linear transformation, and the EU representation is cardinal in this sense. However, the EU formula (1) above does not ensure by itself that the U functional represents preference differences in the sense of (1*). Bouyssou and Vansnick (1990) analyse in detail the formal relationship holding between the classical cardinalist ('difference in preference') and the vNM ('preference among risky prospects') functions.

[70]But see Pareto (1906) for the most famous exposition.

[71]Cf. in particular Hicks and Allen (1934).

[72]The distinction between risk and uncertainty is classic since Knight (1921): in a 'risky' prospect, given (objective) probabilities are assigned to states of the world; in an 'uncertain' prospect, the probabilities are subjective.

[73]It should be stressed that (as for many of the authors mentioned in this section) vNM's interest in utility theory was self-confessedly merely "opportunistic", as the "treatment [of the conceptual and practical difficulties of the notion of utility] is not among the primary objects of [our] work" (1944, p. 8). A model of decision under risk was needed, in particular, in order to solve games with mixed strategies, but vNM's game theory does not need to rely, strictly speaking, on any particular interpretation of payoff figures. The utility interpretation may have been simply an attempt to capture economists' interest.

shown how a utility function may be determined by observation of choices among risky prospects.[74] The measurement procedure is also implicit in vNM, and has since then become well-known: starting with three outcomes x, y and z such that $x > y > z$, the utility scale is normalized by assigning $U(x) = 1$ and $U(z) = 0$. A subject is then presented with the choice between the prospect y for sure and the prospect $[p, x; (1 - p), z]$; if she prefers either the risky prospect or the sure outcome, the probabilities of the former are varied until she is indifferent between the two. At this stage, the expected utility of y can be computed by $U(y) = p$, and the procedure iterated for any value between x and z. The resulting utility curve will display some interesting features: its shape, in particular, will be representative of the agent's 'attitude towards risk'. If we define a risk-averse subject as one who prefers the expected monetary value of a prospect rather than the prospect itself, and if his utility curve can be represented by a vNM function, then the second derivative of the U function of a risk-averse subject will be negative. Following Pratt (1964) and Arrow (1971), the 'absolute risk-aversion coefficient' $-\frac{U''(x)}{U'(x)}$ is customarily used as a measure of risk-aversion.[75] Importantly, the use of derivatives in this formula, and the whole discussion of risk attitudes in the vNM theory, assumes that it makes sense to compare differences in utility values. Hence, it assumes a notion of cardinalism - albeit not the same as 'classical cardinalism'. The latter - Bernoulli's - belongs to the theory of riskless choice, whereas the vNM notion is specific to the theory of risky choice. Pareto's and his follower's 'ordinalism' is explicitly directed against 'classical cardinalism' but has arguably no negative impact on 'vNM cardinalism'.[76]

The importance of vNM's book lies in its providing the first formalisation of the theory, although the goal of a complete axiomatisation was achieved only later with Marschak (1950), Herstein and Milnor (1953), Luce and Raiffa (1957) and other works. A *representation theorem* for (1) was proved from the following axioms:

---

[74]Ramsey (1926) devised a general method for 'measuring' utilities and *subjective* probabilities; his approach to uncertainty was to be pursued further by Savage (1954).

[75]The absolute magnitude of the second derivative of the expected utility function, in fact, cannot be used as a measure of risk-aversion because like expected utility functionals it is unique only up to a linear transformation. Hence the ratio formula adopted by Pratt. For a discussion of some methodological problems arising from the Arrow-Pratt measure, see Hansson (1988).

[76]A majority of today's theorists working in the tradition of 'ordinalism' accept the vNM notion of cardinality as being unproblematic, while they strongly reject Bernoulli's. See the update accounts of the 'two cardinalities' issue in Bouyssou and Vansnick (1990) and Fishburn (1989).

(A1) $>$ is a weak-order relation:

$(x > y) \Rightarrow \neg(y > x)$ [asymmetry]

$(x > y \;\&\; y > z) \Rightarrow (x > z)$ [transitivity]

(A2) $(x > y > z) \Leftrightarrow [px + (1 - p)z > y > qy + (1 - q)z\,]$

for some p and q strictly between 0 and 1 [continuity]

(A3) $\forall p$ such that $0 < p \leq 1$,

$(x > y) \Leftrightarrow [px + (1 - p)z] > [py + (1 - p)z]$ [independence].[77]

Formally, the theorem states that if an ordering $>$ satisfies (A1), (A2), (A3), there exists a real utility function U (defined on outcomes) such that: for every two lotteries x and y,

(3) $x > y \Leftrightarrow EU(x) > EU(y)$,

where EU is defined as in (1) above. Furthermore, the class of U' which satisfy the representation condition (3) is exactly the class of positive affine transformations of U.

The proof of a representation theorem is a very important step in the history of EU theory. Firstly, the axioms stated the restrictions imposed on individual preferences in order to obtain an EU function in a clear and rigorous fashion, and built a fixed target for possible counterexamples - a point towards which I shall turn my attention in the next section. Secondly, *with the proof of a representation theorem rationality came to be regarded as implicitly defined by the axioms.* This is a crucial point: the first formulation of a rationality principle was achieved as a result of the solution to the problem of operationalising utilities. In the early fifties, the problem of rationality got intermingled with the problem of operationalisation: the following meaning-shifts of the rationality principle and later attempts to find a justification to its different formulations cannot be understood without taking such a background into account. Third, in a loop of justification, the *prima facie* 'reasonableness' of the axioms provided the empirical hypothesis that economic agents are EU maximizers (put forward, e.g., by Marschak, 1950) with some plausibility, thus shifting the burden of proof towards the (at this stage, still incipient) opposition.

---

[77]Here the *strict* preference relation '>' is chosen as the primitive concept. Alternative, but equivalent, axiomatisations can be given in terms of the *weak* preference relation '≥' ('is at least as preferred as').

At the end of 1952, Allais produced a series of empirical counterexamples to EU theory. In a famous and controversial experiment (from then on known as 'the Allais paradox'), he asked people to choose first between a lottery (a_1) with the sure prize of one million dollars and a lottery (a_2) with a 10% chance of winning five millions, an 89% chance of winning one million, and a 1% chance of winning nothing. Then, he asked them to choose between a_3, a lottery with a 10% chance of winning five millions and a 90% of winning nothing, and a_4, a lottery with an 11% chance of winning one million and an 89% chance of winning nothing. The gambles are represented in the decision matrix below, with s_1, s_2, s_3 for possible states of the world with the respective probabilities attached to them.

	s_1 (p = 0.10)	s_2 (p = 0.01)	s_3 (p = 0.89)
a_1	1,000,000	1,000,000	1,000,000
a_2	5,000,000	0	1,000,000
a_3	5,000,000	0	0
a_4	1,000,000	1,000,000	0

A majority of the subjects chose a_1 in the first experiment, and a_3 in the second, in apparent contrast with EU theory predictions, and in particular with axiom (A3). According to the latter, the outcomes under s_3 should not be relevant for the choices in the first $\{a_1, a_2\}$ and in the second $\{a_3, a_4\}$ experiment; therefore EU theory prescribes treating the two choices as identical. Allais' primary goal was to demonstrate the superiority of his own account of risky choice over the EU approach. He endorsed classical cardinalism as I defined it, and thus assumed the existence of a cardinal utility function $u(.)$ derivable from agent's choices under certainty. Taking $u(.)$ to denote a classical cardinal function satisfying (1*), and $U(.)$ to denote a vNM utility function as measured by observation of choices among risky prospects, Allais intended to deny that $u(.)$ and $U(.)$ are equivalent (that is, that $u(.) = U(.)$, up to a linear transformation).

Allais' belief in classical cardinalism was supported by evidence independent of risky choices, in particular by introspection, along the lines of Jevons and Marshall. Referring to Savage's defence of the equivalence above, Allais claimed that "there is no justification for *using the term 'index of utility' to refer to the index whose existence he demonstrates subject to certain conditions.* It is impossible to show that this index represents cardinal utility in the usual sense of the term" (1979, p. 509). Before Allais' challenge Friedman and Savage (1948)[78] and Baumol (1951) had interpreted vNM's 'risky' U(.) function as equivalent to the classical cardinalist 'riskless' u(.) function. This was surely the 'natural' interpretation, given vNM's rather unclear statements on the subject matter.[79] Samuelson (1950), Weldon (1950), and Arrow (1951), on the contrary, had argued that the new vNM-utility should have been kept distinct from old classical cardinalist utility. Allais' critiques were crucial in making EU theorists aware of the distinction among the two concepts, a distinction further highlighted by Ellsberg (1954), and then by Luce and Raiffa (1957) and Baumol (1958). After 1952, orthodox decision theorists generally endorsed the 'two cardinalities' position, although accidental identifications of U with u never ceased to be made, and a handful of theoretical dissenters remain.[80] Allais, on the opposite wing, kept arguing for the 'one cardinality' position, but identifying the true cardinal utility with the classical difference-in-preference notion.

<table>
<tr><th></th><th>$u(.) \cong U(.)$</th><th>$u(.) \neq U(.)$</th></tr>
<tr><td>one cardinality</td><td>Friedman & Savage (1948), Harsanyi (1977)</td><td>Weldon (1950), Baumol (1951), Allais (1953)</td></tr>
<tr><td>two cardinalities</td><td colspan="2">Samuelson (1950), Arrow (1951),
Friedman & Savage (1952), Savage (1954),
Ellsberg (1954),
Luce and Raiffa (1957), Baumol (1958)</td></tr>
</table>

Allais' strategy seems the more remarkable, because the diagnosis of failure of the EU hypothesis in his experiment is in fact logically independent of his conception of utility.[81] Let us denote these two elements of Allais' work by (i) - the proposal to keep a riskless utility function as the fundamental tool of analysis

[78]But see their rectification in Friedman and Savage (1952).

[79]Regarding the position of vNM (1944), see Ellsberg (1954) and especially Fishburn (1989), who has shown how the distinction between the two cardinalities can be detected in the footnotes of the *Theory of Games*.

[80]Harsanyi (1977) is probably the most vigorous (but isolated) defender of this standpoint among today's decision theorists, but it is unclear whether he adheres to it for reasons other than the fact it is needed for his defense of utilitarian ethics. See Mongin and D'Aspremont (1996).

[81]For a retrospective outline of Allais' strategy, see his (1992).

for risky decisions as well, and (ii) the falsification of the EU hypothesis, respectively. When, in 1952 during the C.N.R.S. Colloquium in Paris, Savage accepted that (i) $u(x) \neq U(x)$, Allais concluded that (ii) "as a matter of fact, the neo-Bernoullian index whose existence Savage claims to have proved *only exists on the paper*" (1979, p. 509). In fact, not only (i) "to treat this index on the same footing as cardinal utility is *deliberately to delude* the unwarned reader into believing that the theory on which it is based provides a measure of [classical] cardinal utility" (ibid., p. 509), but (ii) the function $U(x)$ could only represent the function maximized by an *ideal* agent obeying the vNM axioms, and unfortunately this does not seem to be the case for *real* agents.

In other words, the Allais paradox can be interpreted in at least two ways. On the one hand, (i) moving from the assumption that agents maximize a classical cardinal utility function, one can use the empirical data to argue that it is impossible to measure it by means of vNM's method. The difference in value [u(1,000,000) - u(5,000,000)] as measured in a certainty context cannot be used to predict the outcome of the risky experiment. On the other hand, (ii) it is possible to use the data to show that the preferences displayed by subjects are inconsistent with EU axioms, and therefore that no 'index' of preferences $U(.)$ can be constructed from them. If preferences were consistent with the independence axiom, in fact, there would exist a function linear in the probabilities representing such preferences, contrary to experimental evidence.

Although the tradition has mainly retained aspect (ii) of Allais' work - i.e. the falsification of the EU hypothesis - it should be stressed that it is just *one* aspect of his rich and sophisticated approach. The thesis (i) about the existence of a riskless cardinal utility function and of its primacy over risky indexes of choice, despite its being logically independent of (ii), is crucial in order to understand Allais' positive heuristics in all multifarious aspects, and also the reasoning behind his experimental achievements.

2.3. Is every 'counterexample' a counterexample? Normativism and its problems

Before the Paris conference, EU theory had been interpreted mainly as a descriptive theory of human behaviour. The Allais paradox provoked a major shift in the problem of decision under risk, and many decision theorists began to defend EU theory from a normative point of view. The roots of economics being in the field of the moral sciences, normative interpretations are much older than any 'descriptivist' view. 'Normativists'[82] claim that their theories aim to describe the behaviour of an ideal rational agent, but not only because such an 'ideal type' is useful in order to understand the behaviour of real agents; rather, models of rational behaviour are significant for evaluative purposes.[83]

The 'Neo-Bernoullians' - as Allais labelled the strong defenders of EU theory - slowly shifted to a different terrain in order to save their model of rational choice: theories of rational behaviour have a normative status and, therefore, should not be modified in the light of 'irrational' choices. According to Allais, the shift from a descriptive to a normative interpretation deprived EU theory of its scientific content. Nevertheless, Allais accepted the challenge and, as we shall see, gave a substantial contribution to the subsequent debate about normativism. This was an extremely important move: had Allais decided not to fight his battle on both fronts, much of the methodological debate I am going to reconstruct would have simply not have taken place. The clash focused on the meaning and content of the rationality principle, and led to a series of meaning-shifts which proved to be very fruitful for the development of decision theory. In the next sections I shall try to assess these shifts and analyse their theoretical implications.

It is interesting to notice that, despite his being primarily interested in the descriptive value of EU theory, Allais in a sense was forced to take the normativist position at face value. Marschak (1950) had argued that the *prima facie* 'reasonableness' of the axioms confirmed the validity of the model as a 'first approximation' to human real behaviour under risk, thus elaborating on vNM's standpoint.[84] Real economic behaviour, as de Finetti (1952) had put it, can be seen as the result in the aggregate of rational behaviour plus an error component. By devising a model of rational behaviour, therefore, progress both on the descriptive and the prescriptive dimensions is achieved. Allais similarly subscribed to a heuristics of successive approximations to a true, complete and accurate description of human decision under risk, but followed a different strategy. He decided to start by modelling an introspectively well-established human motive, i.e. the desire to maximize 'absolute satisfaction' (a psychophysical sensation in the Fechner-Weber sense), and then add other factors arising from the perception of risk. Real human behaviour, to him, is the result of the attempt to maximize utility, plus the impact of four 'fundamental factors' and of ten 'secondary

[82]Of course 'normativism' and 'descriptivism' are labels created to better illustrate my point, and do not correspond to any real 'school' in economics or decision theory.

[83]In modern economics, the origins of this position can be traced back to Walras.

[84]Cf. also his remarks in C.N.R.S. (1953), pp. 25-26, 35, 152.

factors'. The results of his experiments, for example, could not be explained by EU theory because the latter neglects a fundamental factor (the 'dispersion of psychological values') which is present in almost all risky choice-situations.[85] Without going into the details of Allais' own theory, it is important to stress that in order to hit the 'Neo Bernoullian' position at its heart it was necessary to show that the axioms of EU theory are not so reasonable after all. This way, even the 'heuristic value' of vNM's approach could be denied.

As noted, the first normativist interpretation of vNM's model was probably due to Marschak (1950), who also proposed to take the theory as descriptively valid to a 'first approximation'.[86] Arrow (1951) and Baumol (1951) followed the same approach. Savage, who had in the papers written with Friedman put the theory to descriptive use, focused in later papers on its normative properties. In (1951), (1952), and then in his *Foundations of Statistics* (1954), subjective EU axioms are presented and defended as describing the behaviour of a rational agent facing risk and uncertainty. Savage's and Allais' positions constituted the opposite poles of the debate that took place at at the Paris Colloquium of 1952. Savage (1952) presented his subjective EU theory explicitely as a normative theory of behaviour from the start. He acknowledged that *sometimes* people's behaviour should agree with the theory, which suggests a possible empirical test 'in certain cases' (1952, p. 29), but his emphasis is very different from that of the Friedman and Savage articles (1948; 1952). Allais, in contrast, challenged Savage's model as a theory of rationality, by arguing that certain effects (like the 'complementarity effects' at the origins of his own paradox) cannot be deemed 'irrational' (cf. C.N.R.S., 1953, p. 39).

Among the other participants to the 1952 conference, de Finetti (1952) was also keen on defending EU theory from a normative standpoint. (This was probably due to his commitment to Bayesianism as an interpretation of subjective probability.) Like Walras before him, de Finetti argued that, in case of deviations, agents should be led by correction to approximate the rational model (in C.N.R.S., 1952, pp. 52-53). Samuelson's position was most curious: in (1950) he felt neutral on the issue for lack of convincing arguments. In 1950 Savage introduced him to Dutch-Book arguments and convinced him of the normative validity of EU theory. Still, he remained sceptical of the adequacy of the model to

[85] Cf. Allais (1953/1979), pp. 39-56.

[86] Marschak's twist was however consistent with the spirit of the time: research with strong normative connotations was carried on in the forties in a number of centres (such as the Cowles commision) and in operations research in general.

represent one-shot choices, when there is no possibility of exploitation or learning.[87] According to Samuelson (1952), the main virtues of the vNM axiomatization are ‘aesthetic and semantic’: the model is in continuity with the coherentist approach to preferences that became orthodox with the ‘ordinalist’ revolution in consumer theory. Yet, Samuelson during the conference did not refrain from defending the ‘reasonableness’ of his own Strong Independence Axiom from Allais’ criticism (cf. C.N.R.S., 1953, pp. 147, 157, 163).

The main problem of the ‘Neo-Bernoullians’ was to ‘neutralise’ the anomalous evidence and to deny it the status of counterexample to EU theory. By endorsing the normative standpoint, they could reject all the apparent counter-evidence as due to 'mistakes', 'errors of calculations', etc. Imre Lakatos has described defensive strategies of this kind - which he labelled 'monster-barring' (or 'monster-elimination') - in his *Proofs and Refutations* (1963-64/1976), a long essay devoted to the logic of mathematical discovery. This procedure consists in narrowing the domain of a theory through 'ad-hoc' redefinitions of its terms which cut off dangerous falsifiers. In his case study, Lakatos showed how mathematicians rejected counterexamples to Euler's conjecture by simply denying that a non-Eulerian polyhedron is a ‘real’ polyhedron. In our case, normativists tried to reject counterexamples claiming that EU theory is a theory of rational behaviour and non-vNM behaviour is not ‘really’ rational. Much of the debate between Allais and his normativist collegues focused on the interpretation of normativism, on the legitimacy of this defensive strategy, and on the distinction between irrationalities and genuine counterexamples.

How can a normative theory be confirmed or refuted? Such a question lies hidden behind much of the early controversy on EU theory. At the beginning, no general agreement existed about a methodology for assessing normative theories. Several different (sometimes only slightly different) points of view can be identified; other, more sophisticated, ones slowly emerged as the controversy progressed.

At the 1952 Colloquium, Jacob Marschak held that "des postulats choisis avec bon sens imposent naturellement des restrictions aux fonctions [...] admissibles" (in C.N.R.S., 1953, p. 153). The notion of *bon sens*, unfortunately, is an intrinsically vague one: *whose* ‘good sense’, to begin with? Suppose that, as

[87]Cf. the 1965 postscript to Samuelson (1950), in *Collected Papers*, vol. 1, pp. 124-126.

in the case we are concerned with, there exists a radical disagreement about the reasonableness of some principle: how can we resolve it?

The most drastic solution to such problems would be a sort of ‘postulationism’: to define rationality as ‘obeying the EU axioms’. But such a move would trivialize the notions of 'making a mistake' and 'irrational behaviour' that - correctly, from an intuitive point of view - should demarcate admissible from inadmissible counterexamples. It becomes impossible to distinguish 'making a mistake' from 'not obeying the EU axioms'.[88] The latter being just a definition of 'counterexample to EU theory', all counterexamples are according to this perspective instances of mistakes, irrational behaviour, and thus devoid of significance. It is not surprising then, that nobody seriously pursued the strategy.[89]

Postulationism, thus, seems to be a dead end: why defining rationality in terms of the EU axioms, instead of other, possibly incompatible, principles? As noticed by Paul Samuelson, "de même qu'on ne peut pas discuter de goûts, de même on ne peut pas, du point de vue de la déduction pure, discuter sur les axiomes" (Samuelson, 1952, in C.N.R.S., 1953, p. 143). From a purely formal point of view any definition is as good as any other, and thus postulationism is acceptable only in a more 'liberal' form, which we may call ‘coherentism’. Coherentists argued that the behaviour exhibited in Allais’ experiments can be deemed 'irrational' if and only if the subjects committed themselves to follow EU axioms: "[the] normative property is based on the acceptance of the theory" (Morgenstern, 1972, p. 712). Amihud similarly argued later that

> The theory merely claims that *if* in a certain situation, given a choice between certain lotteries, *an individual agrees that his behavior is in accordance with the von Neumann-Morgenstern axioms - then an inconsistency in choices may be pointed out as irrational.* [...] Clearly, an observation of any pattern of choice whatsoever is insufficient by itself to determine rationality (or irrationality). Irrationality is rather a result of an action taken by an individual who has agreed to accept this particular definition of rationality (Amihud, 1979, pp. 149-150).

---

[88]Notice that there the postulationist position still admits the theoretical distinction between errors derived from the adoption of some ‘wrong’ decision rule, and other trivial mistakes such as calculation errors. In practice, however, the two become indistinguishable.

[89]In his (1979) reply to his critics, Allais tries to attibute such a position to Amihud, but it is clearly a polemical misrepresentation of his position.

From this point of view, a 'mistake' can only be an error in deducing the ‘right’ action from a set of rules that the agent accepts as defining rationality. We now have a way of detecting ‘mistakes’: by asking agents what their own theory of rationality is. If their choices conflict with their theory, then we can claim that they have made a mistake. Irrationality can only be defined as inconsistency between a subject’s theory and his behaviour.[90] Morgenstern and Amihud played a rather marginal role in the EU controversy, but this idea is very important indeed. It lies, as we shall see soon, at the origins of the empirical work done by experimental psychologists on people’s readiness to revise their choices in the light of (either formal or informal) theories of rationality.

Morgenstern and Amihud’s strategy, however, is not very effective in the context of a debate such as that on decision theory. It cannot, in fact, help very much to make progress in a debate aimed at finding a *unique* rationality principle. Coherentism leaves anyone free to choose whatever principle of normative choice she wants to endorse. Why should someone hold a principle of rationality rather than another? As Allais pointed out, this amounts merely to saying that "*to act as a normative guide a theory must be accepted as valid*", "*but*", of course, "*precisely the question is to know which theory* [...] *should be adopted*" (1979, p. 544-545). We have not made much progress, and still have no clue of what it means for the agent’s *behaviour itself* to be coherent.

Of course, we all have a rough idea of what a rational choice is. Formal models may indeed be seen as (fallible) attempts to turn our intuitions into a precise, rigorous formulation. In this sense, the informal notions of rationality may play a regulative role in assessing formal models. The latter - unless we are certain of a perfect overlapping - must be corrigible in the light of some counterexample. To a fallibilist-normativist, the problem of the legitimacy of counterexample changes, so to speak, into the problem of ‘monster-acceptance’. Counterexamples are useful for theory-improvement, but not every counterexample should be allowed to falsify a theory.

The problem of the nature of counterexamples is particularly puzzling in the case of normative theories. In the case of empirical theories, falsifiers should in the

[90] Notice the contrast between such a point of view and the naive idea that there can be mere behaviour-behaviour inconsistencies: nothing can be inferred about the rationality of my choices from the mere fact that I choose something in one situation and something else in another - that is, without some theory of rationality enabling to interpret my behaviour.

best case[91] point to the falsity of the universal laws under test. According to Lakatos (1967), falsifiers in mathematics are 'heuristic' ones, they point to the inadequacy of a formal system to 'catch' an underlying informal theory.

> If we accept the view that a formal axiomatic theory implicitly defines its subject-matter, then there would be no [...] falsifiers except the logical ones. But if we insist that a formal theory should be the formalization of some informal theory, then a formal theory may be said to be 'refuted' if one of its theorems is negated by the corresponding theorem of the informal theory (Lakatos, 1963-64/1976, p. 36).

These ideas seem to fit very well the case we are concerned with. In 1953 Allais proposed a 'quasi-empirical' criterion in order to 'catch' an informal minimal principle of rationality:

> Rationality can be defined by having regard to the behaviour of persons who are commonly considered rational.[92]

But people can be legitimately considered rational if and only if "*there exists independent grounds, i.e. without any considerations of random choices, for believing that they behave rationally*" (Allais, 1979, p. 467). That is, an *abstract* criterion sustains the experimental, because "only those whose behaviour is in line with the first are usually considered as rational" (ibid. p. 467). Agents acting in situations of certainty in accordance to the first two axioms are rational, in that they obey to the abstract minimal definition of rationality as consistency.

> A man will be deemed to act rationally
> (a) if he pursues ends that are mutually consistent (i.e. not contradictory)
> (b) if he employs means that are appropriate for these ends (Allais, 1953/1979, p. 78).

Moreover, satisfaction of (first order) stochastic dominance should be taken as a necessary prerequisite of any normative theory of decision.[93] Such an

---

[91]The qualification is needed because of the Duhem-Quine problem (see the next chapter).
[92]See for instance Allais (1953/1979), pp. 80, 86; (1979), p. 467.
[93]Allais called stochastic dominance 'the axiom of abslolute preference', and made it part of his definition of rationality (cf. Allais, 1953/1979, pp. 78-80). A lottery *X* (first-order) stochastically dominates *Y* whenever the probability assigned to the preferred outcomes is higher in *X* than in *Y*. In other words, the cumulative distribution of *X* is *below* the cumulative distribution of *Y*.

abstract, minimal and allegedly universally accepted definition of rationality as consistency at the same time provides a criterion for appraising counterexamples, and 'cuts off' those axioms which (because they are not implied by (a) and (b) above, e.g. independence) are open to revision in the light of counterexamples. Allais allowed counterexamples such as those emerging from his own experiments to act as normative falsifiers by 'stretching' the concept of rationality: the latter is now identified with a broader notion only partly overlapping with the axioms of EU theory.[94]

In a late reflection on the normative status of his own theory, Oskar Morgenstern argued in a similar vein that "a norm follows only from [...] another norm of more general content", and that "this process should ultimately lead to some original norm, which is simply given [taken as] and not derived from any other" (1972, pp. 710-711). The problem then is: at what level will agreement be reached regarding an acceptable general rationality principle?

Unfortunately, Allais' criterion of acceptance of falsifiers is still too weak. It solved the problem Allais wanted to solve, i.e. it provided a criterion liberal enough to legitimise all the counterexamples known at the time - but only those counterexamples that hit the independence axioms. The criterion, to put it in a different way, is rigid because strictly linked to a specific definition of rationality: more formally, it is modelled on the notions of preordering (axiom A1 above), and of stochastic dominance. But why accept these two principles? The procedure of appraisal of counterexamples had to be further refined.

A crucial step forward in the debate was made when Savage (1954) proposed his 'quasi-empirical' test of rationality while debating with Allais. It is, I think, the best formulation of a critical method of theory-improvement to be found in the debate on EU theory. According to Savage's quasi-empirical test of rationality,

---

[94] 'Concept stretching', according to Lakatos is the other side of 'monster-barring'. "For any proposition there is always some sufficiently narrow interpretation of its terms, such that it turns out true, and some sufficiently wide interpretation such that it turns out false. Which interpretation is intended and which unintended depends of course on our intentions. The first interpretation may be called the *dogmatist, verificationist or justificationist interpretation*, the second the *sceptical, critical or refutationist interpretation* (Lakatos, 1963-64/1976, p. 99). On 'inflation' and 'deflation' of meaning and Lakatos' dialectical logic see Larvor (1998), ch. 2.

> If, after thorough deliberation, anyone maintains a pair of distinct preferences that are in conflict with the sure-thing principle, he must abandon, or modify, the principle; for that kind of discrepancy seems intolerable in a normative theory. [...] In general, a person who has tentatively accepted a normative theory must conscientiously study situations in which the theory seems to lead him astray; he must decide for each by reflection - deduction will typically be of little relevance - whether to retain his initial impression of the situation or to accept the implications of the theory for it (1954, pp. 101-103).

During the 1952 Colloquium in Paris, Allais presented Savage with a version of his questionnaire, and Savage's choices, in line with the majority of people, violated vNM's axioms. Thi story has since then become part of the folklore (Savage was at the time working on his *Foundations of Statistics*!). Savage later revised his answers, recognising that he had been 'irrational', and that he would have rather followed EU prescriptions. His quasi-empirical test was conceived in order to justify this move. In the passage quoted above, Savage - still holding normativism - recognized that EU theory (and the independence axiom in particular) was in principle revisable in the light of counterexamples. Then, he identified normativity with the *convincing power* of a theory, and proposed his test. Although his method was a step forward in the debate, because it was not founded on any formal (however minimal) definition of rationality, it still needed a decisive improvement. In particular, the ambiguity of Savage's reference to "thorough deliberation", "conscientious study of situations", "reflection", made his method apparently almost empty: how does one know whether he should "retain the initial impression" (that is, accept the counterexample) or "accept the implications of the theory"?

Allais noticed that Savage's method was questionable if tendentiously applied: "Savage presents my counterexample without saying a word about the general theory on which it is based" (1979, p. 535). When seen 'through the spectacles' of EU theory, of course "*the* [counter-]*example which is only a specific illustration of a general theory, is pinpointed as a wayward curiosity concealing an 'error' which is easily brought to light*" (Allais, 1979, p. 537, emphasis in the original).

Allais' remarks highlighted the trivial but crucial point that the alternative is not among counterexamples being 'rational' or 'irrational' *per se*, but 'rational' or 'irrational' relatively to some theory of rationality. Before asking whether agents want to revise their choices or not, *both* rival theories should be proposed.[95] This suggests that the problems of criticism and growth of (normative) knowledge cannot be separated. The problem of 'theory-choice' cannot be solved independently of the problem of 'monster-acceptance', because we need counterexamples in order to falsify and improve a theory of rational behaviour. On the other hand, the problem of 'monster-acceptance' cannot be solved independently of the problem of 'theory-choice', because only a new conjecture about the rationale of the falsifying evidence can legitimate a monster and transform it into a counterexample.

Again, the problem of the nature of normative falsifiers emerges in its complexity: counterexamples are tricky because we do not have intuitions about them. This is a typical feature of normative theories, which distinguishes them from empirical ones: counterexamples are disturbing because we do not know what our attitude towards them ought to be. Intuition is not the ultimate criterion of appraisal. It must be guided by reasons, and shaped by arguments. This applies to counterexamples in both their interpretations. We need an argument in order turn a counterexample into a 'monster' as much as to turn it into a 'heuristic falsifier'. This feature plays a very important role in the dialectics of the debate, because monsters stimulate proofs in order to justify their rejection.[96]

To sum up: according to the Savage-Allais method, some intuitive, but so far formally undefined, principles of rational behaviour exist. These informal principles regulate the acceptance of counterexamples. So, it is crucial that we have at least an intuitive idea of the factors which may have provoked the emergence of a counterexample, and moreover that these factors cannot be deemed irrational in the light of the informal rationality principles which guide agents' choices. Otherwise, a 'monster-barring' defense becomes legitimate and the counterexample can be labelled as an instance of irrational behaviour. Implicitly, the Savage-Allais method will accept as superseding theories only models of

---

[95] Allais did not seem to get this point, for instance when he claimed (consistently with his own methodology presented above) that his experiments proved "that there are people who are considered as rational and who take decisions that are incompatible with this [i.e., vNM's] formulation. That is enough for me" (1979, p. 546). As I am trying to argue, I think that it is *not* enough.

[96] "*Intuitionism is here used not for providing foundations but for providing falsifiers, not for discouraging but for encouraging and criticizing speculation!*" (Lakatos, 1967, p. 38).

rational choice: behind a counterexample there must be a rival theory, and behind a rival theory there must always be a rationality principle.[97]

This method is liberal enough to allow the emergence of counterexamples, but strict enough not to let every counterexample count as a falsifier. The intuitively plausible idea of human 'mistakes' is preserved and plays a role in the heuristic strategy. But is the method too strict? Does it not require that a new, better theory be available before we take the old one as falsified? Morgenstern (1979) argued in such a methodological-falsificationist vein that Allais produced a series of counterexamples without a superseding theory, and therefore his refutations could not be taken seriously.[98] But of course we should distinguish rejection from falsification: in order to consider a (normative) theory falsified, we do not need a fully-developed rival theory, a promise is enough. In contrast, an isolated counterexample is *not* enough: we need at least a 'sketch' of a new theory, a rough conjecture about the origins of the anomaly. As we shall see, in the light of the informal principle of rationality, the clash between the anomaly and the former model of rational decision should in the best cases also provide heuristic advice about how to develop a new theory that might supersede the old conjecture.

Allais is then responsible for at least three distinctive methodological contributions to the debate: a refining of the idea of intuition as imitation of the ‘rational man’; his well-defined but less demanding notion of rationality; the sophisticated falsificationist view that counterexamples come with alternative theories in support. From a scientific point of view, Allais provided the sketch of a new theory, but thirty years went by before mathematically acceptable superseding models were put forward by Mark Machina, John Quiggin and others. Before then experimenters had already shown that a majority of subjects, when provided with informal arguments, refuse to revise their ‘paradoxical’ answers to Allais’ test. Agents’ intuitive rationality appears stronger than economists’ formalisations. MacCrimmon (1968) and Slovic and Tversky (1974) tried in a number of experiments to implement Savage’s and Allais’ suggestions, thus creating some real quasi-empirical tests of rationality. The experiments confirmed Allais’ intuition - that the quality and variety of the arguments (the

---

[97] Sen’s (1985) notion of ‘reflection rationality’ captures some of the features of the Allais-Savage quasi-empirical methodology. Sen fails, however, to notice that it is a ‘meta-level’ regulative principle.

[98] Such a claim is now outdated, thanks to the developments in decision theory to be mentioned in next section.

'theories of rationality') provided by experimenters influence crucially agents' decisions whether to revise their choices in Allais-type experiments or not.

2.4. Generalisations

Up to now I have focused on the debate concerning the normative status of EU theory and on its crucial role in making the anomalies acceptable as genuine falsifying evidence. To accept such evidence does however open the problem of how to revise the refuted theory and replace it with a better one. In this section I shall show that normative reasoning played another important role in helping to solve this particular problem. The 'minimal principles' that emerged from the debate on rationality were used as heuristic tools for the discovery of superseding models of rational choice under risk.

Let us, to begin with, sketch a very broad heuristic principle common to decision theory and perhaps economics in general, stating that every new theory should be a theory of rational individual behaviour. Such a principle is compatible with other (more specific) ones: Allais' idea that one should build on the classical utilitarian tradition and identify $U(.)$ with $u(.)$, for example; or the commitment to operationalism that was common to several economists at the time. According to the general principle, at any rate, only rational models of decision will be acceptable as superseding ones. Adding more detail, we shall see that the development of alternative models of decision proceded in two steps. The first one was mainly dictated by Allais' minimal principle of rationality, telling decision theorists which of the axioms of EU theory to abandon and which ones to keep. Such a move led to the development of weaker models, able to account for the refuting evidence but at the same time with less empirical content than the original one. The second step consisted in deriving stronger models from the weak ones and some of the anomalous evidence.

Ideally, a full reconstruction of a cognitive process of theoretical discovery could identify a rigorous deductive argument from a set of meta-principles (metaphysics, methodological rules, tacit commitments of all sorts), the empirical evidence known at the time, and the older refuted theory, to the new rival theory.[99] In practice this is often difficult to effect, because most of the reasoning lies hidden in scientists' mind, and I shall certainly fall short of this goal. I shall however present a partial reconstruction supporting a non-trivial thesis: that

[99] For some examples of full reconstruction, see Zahar (1983).

principles of rationality were among the premisses of the reasoning that led to the generalisations of EU theory put forward in the early eighties. This will shed further light on the importance of the debate on normativity in the development of decision theory.

An attempt to solve the Allais paradox was provided by Mark Machina's (1982; 1983) 'generalised expected utility analysis'. Like Allais, Machina spotted (A3) as the weak axiom in EU theory, and tried to replace it taking the known counterexamples into account.[100] According to Mongin (1988a, p. 319), the decision to hit the third axiom of EU theory can be reconstructed as a rational move dictated by the mathematical structure of the refuted theory. There exists, in fact, an implicit hierarchy among the axioms of EU theory: the axioms of continuity (A2) and independence (A3) become ‘inefficiently precise’ when imposed on a non-well-ordered relation (A1). Similarly, independence makes little sense without continuity, a necessary condition for proving the existence of a utility function. In this sense, one may say that (A3) ‘implies’ (A2), and both ‘imply’ (A1). The relation of ‘implication’ (figure 1a below) can be reversed to represent the ‘hard core’ and the ‘protective belt’ of vNM’s theory (figure 1b below). In case of falsifying evidence, the ‘arrow of refutation’ will be naturally directed towards (A3) first.[101]

---

[100] Allais has denied that there is any continuity between his and Machina’s work, and started a harsh controversy in his (1988b); see the issue of *Theory and Decision* dedicated to that debate (vol. 38, no. 3, 1995). The received interpretation among decision theorists today is that Machina has captured some important features of Allais’ informal ideas. Allais’ ideas are also at the origin of other generalised models of decision, like Quiggin’s ‘Expected Utility with Rank-Dependent Probabilities’ (1982), which is presented briefly below.

[101] On the notions of ‘hard core’ and ‘protective belt’, already implicit in Lakatos (1963-64/1976), see in particular Lakatos (1970).

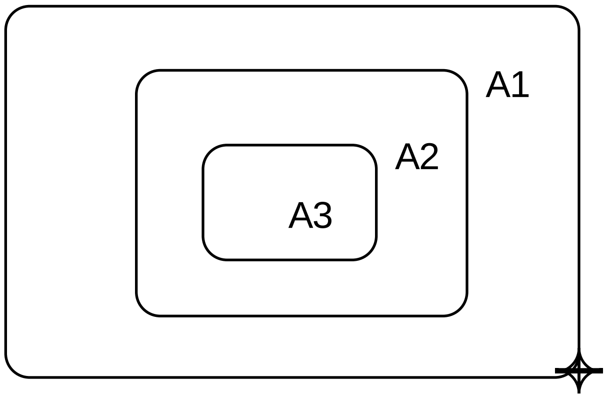

Figure 1a

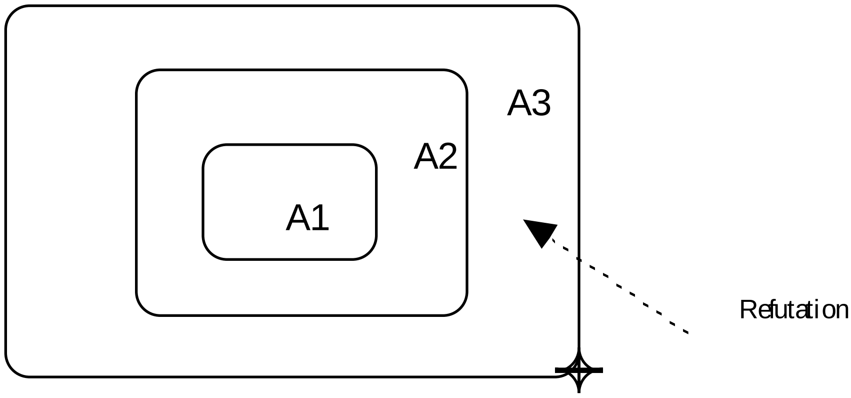

Figure 1b

It is not clear, however, that such considerations *alone* determined the decision to revise EU theory by giving up independence. The reconstruction put forward in the previous sections has shown that the debate on normativism probably played a parallel role, leading to the identification of a ‘normative hard core’ in the axioms of weak-ordering and the principle of stochastic dominance. The minimal notion of rationality proposed by Allais (1953/1979), in other words, explicitly offered the third axiom of EU theory as the ‘natural’ victim of refutation. Generalised models of decision such as Machina’s became acceptable as economic theories due to their being consistent with some (however minimal) principle of rationality. Any historical account of the rise (Fishburn and Wakker, 1995) and fall (Mongin, 1988a) of the independence principle that sharply separates the debate about the descriptive from the debate about the normative virtues of EU theory is bound to miss an important aspect of the story.

Mark Machina presents his positive strategy of revision of the received model in several stages: starting from the familiar point that the independence axiom plays a crucial role in defining a utility function 'linear in the probabilities', Machina showed that most counterexamples are compatible with the weaker assumption that utility is 'smooth (or differentiable) in the probabilities' (Hypothesis 1).[102] Given that linearity is a special case of differentiability, the vNM function becomes a special case of 'generalised expected utility analysis'. However, Hypothesis 1 is incompatible with the suggestion that agents act as if maximizing a unique EU function independently of the prospects they compare. In Machina’s theory, preferences define a set of ‘local utility functions’ maximised by agents when choosing in the neighbourhood of a certain prospect. Local utility functions are increasing in money, thus formalising Allais’ normative intuition that stochastically dominating prospects should be preferred to stochastically dominated ones.

Given that the new theory is apparently a weaker version of the old one, Machina put forward a further restriction in order to fully explain the trend of the violations of the independence axioms, and proposed (Hypothesis 2) that the concavity of a local utility function (and therefore aversion to risk[103]) be related to the property of stochastic dominance: for any two local utility functions over lotteries, if one lottery stochastically dominates the other, the function determined by the dominant lottery must show a greater risk aversion than the function

[102]See Machina (1983), pp. 267-276.

[103]In a formal sense taken from vNM theory, see above section 2.2.

determined by the dominated one.[104] Such a procedure is quite typical of theoretical 'discoveries' in science in general: firstly, the falsified model is weakened by relaxing some of its assumptions; then, the generalised model is made more precise by imposing some new hypothesis suggested by theoretical considerations or empirical data. I shall refer to such a characteristic sequential 'deflation' and 'inflation' of empirical content as the 'zig-zag' pattern of heuristic development.[105]

Generalised EU analysis provides an interesting example of how falsifying evidence can help to build a new conjecture superseding the old (falsified) one. As a presentational device, it is useful to use the so-called 'Marschak-Machina' triangle.[106] The triangle permits the mapping of indifference curves among lotteries in a diagram delimited by a vertical axis upon which probability values p_3 are represented ($0 \geq p_3 \geq 1$), a horizontal axis with values p_1 ($0 \geq p_1 \geq 1$), and by an hypotenuse obtained by linking $p_1 = 1$ and $p_3 = 1$. We can represent lotteries with (at most) three preordered outcomes ($x_1 < x_2 < x_3$) on the Marschak-Machina triangle: each point in the diagram corresponds in fact to a value for p_1, a value for p_3, and a value for p_2 (since $p_2 = 1 - p_1 - p_3$). The Marschak-Machina triangle permits the representation of indifference curves associated with agents' preferences, by tracing a line linking the set of lotteries among which the agents are indifferent. EU theory provides some predictions about the shape of such curves. In particular, it requires that indifference curves run from south-west to north-east in the direction of increasing preferences and - as a consequence of assuming independence - be straight lines and parallel to each other. The Allais experiment, on the contrary, might suggest that in some cases indifference curves 'fan out' in the triangle without being parallel. The prospects of Allais' experiment - namely a_1 = (1m, 1), a_2 = (5m, .10; 0, .01; 1m, .89), a_3 = (5m, .10; 0, .90), a_4 = (1m, .11; 0, .89) - and the related choices ($a_1 > a_2$, $a_3 > a_4$) correspond to indifference curves as in figure 2.

The relative steepness of indifference curves represents aversion to risk; since a lottery dominates all those positioned below it on its right in the triangle, indifference curves implied by Machina's Hypothesis 2 fan out non linearly towards the outside (the so-called 'fanning out' effect - see figure 3).

---

[104] See Machina (1983), p. 282.

[105] The term is borrowed from Koestler (1959). Contrary to Koestler and following Lakatos, however, it is argued here that such a zig-zag pattern does not result from 'sleepwalking': scientific discovery can be reconstructed as a rigorous and quasi-rational process.

[106] Cf. Marschak (1950) and Machina (1982).

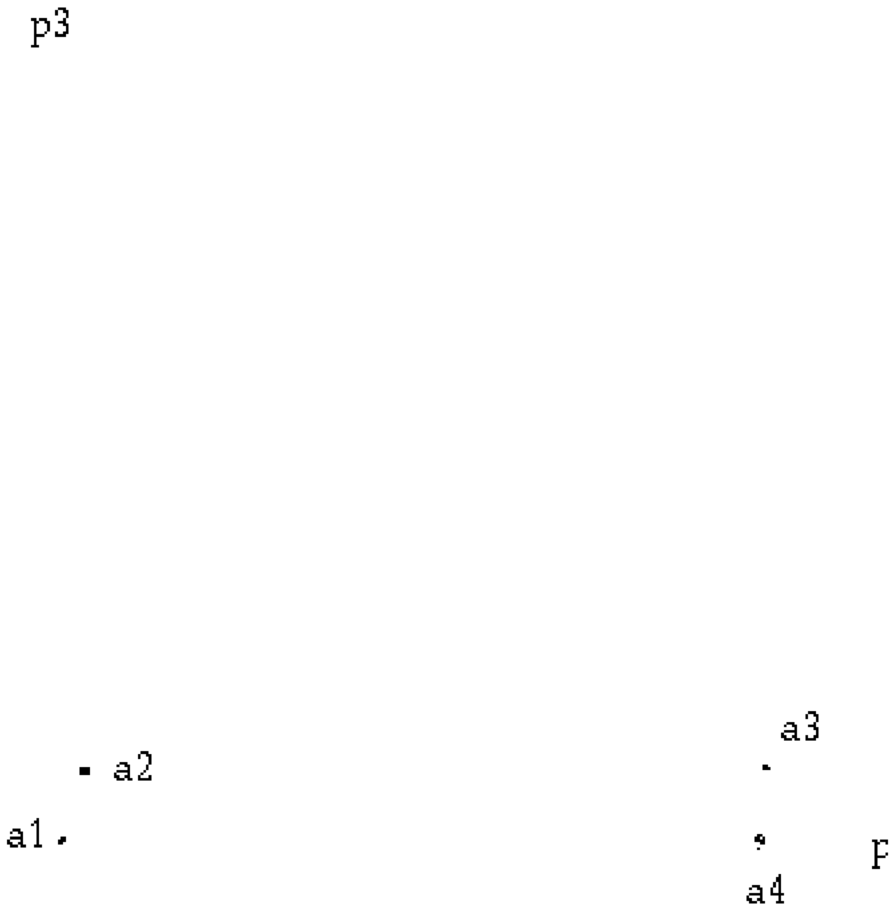

Figure 2

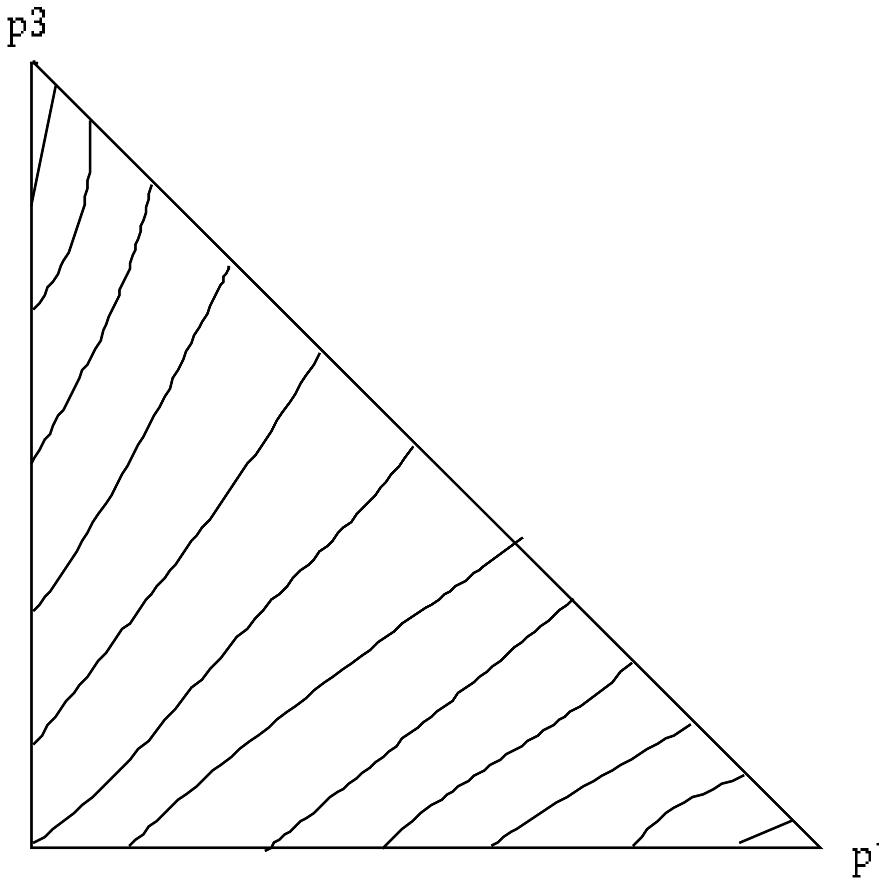

Figure 3

MacCrimmon and Larsson (1979) replicated Allais' experiments and showed that violations of transitivity are more frequent when lotteries have extreme probability values than when lotteries have moderate values. This can be taken to mean that indifference curves near the extremes of the Marschak-Machina triangle fan out significantly, whereas those in the middle are approximately parallel as prescribed by EU theory. Hypothesis 2 essentially builds the falsifying evidence into Machina's theory[107] requiring that the utility functions of a certain lottery exhibit greater risk-aversion than those of the lotteries it stochastically dominates. This procedure highlights once more the crucial role played by the debate about normativism in the development of a descriptive theory of choice under risk. Machina's theory could have never been 'deduced from the phenomena' if such phenomena were treated as mere instances of irrational behaviour. Machina fully accepted Allais' recommendation that the preordering and stochastic dominance preservation assumptions be satisfied, in agreement with the objective of defining a rational choice model alternative to EU theory. Such a step was possible only conditional on a shift at the level of methodology, i.e. in the complicated procedure of acceptance of counterexamples described in the previous sections.

To corroborate this analysis, it is worth comparing Machina's model of rational decision making with a different, purely descriptive, model put forward by the psychologists Kahneman and Tversky (1979) just a few years earlier. Kahnemann and Tversky followed another of Allais' suggestions and tried to model an effect of 'distortion in the perception of the probabilities' allegedly at the origins of Allais' data. The crucial feature of their 'Prospect Theory' is a 'weighting function' $\pi(p)$, which in order to accommodate the known empirical evidence (like the one emerging from the experiments of Allais, Kahneman and Tversky, and others) has to be increasing with, but non-linear in, p. The hypothesis that agents' preferences satisfy first-order stochastic dominance, however, holds if and only if π is linear - that is the special case where Prospect Theory reduces to EU theory. Kahneman and Tversky dealt with this problem conjecturing that dominated alternatives are detected and discarded by subjects prior to the evaluation of the other prospects. Such a solution must have seemed unsatisfactory to a scientific community that increasingly identified progress with formalisation. This may help to explain the rather mild interest raised by Prospect Theory among economists: the theory (in its most elegant version) could not be interpreted as a model of normative decision making.

---

[107] On such a heuristic procedure, cf. Zahar (1983).

Quiggin (1982) and others later devised a theory assigning different weights to different outcomes with the same probabilities, thus building on Kahneman and Tversky's proposal (e.g. 'extreme' consequences get different weights than 'intermediate' consequences, although they all have the same probabilities).[108] Quiggin's new weighting function $h(.)$ is able to satisfy all of Allais' requirements of rationality including the stochastic dominance requirement.[109] Given the fact that today Quiggin's model is probably the preferred generalised model of decision among decision-theoretically oriented economists,[110] we have here a strong confirmation of my thesis: the normativity debate provided meta-criteria of acceptance for superseding theories of choice under risk.[111]

2.5. Intransitive behaviour

I have so far tried to show how the development of contemporary mathematical decision theory was strongly influenced by its ambiguous status, as a descriptive theory of people's actual behaviour and as a normative theory of rational agents' ideal choices. In particular, I have reconstructed the history of how the received

---

[108] Arguably, this theory originates in another part of Allais' work (cf. in particular Allais 1988a). It is also in part due to Yaari (1987) who presented it independenltly but in a more particular form. The so-called 'Rank-Dependent Expected Utility Theory' (or 'Expected Utility with Rank-Dependent Probabilities') quickly established itself as a major alternative to EUT, at the expense of Machina's Generalised Expected Utility Analysis.

[109] Allais, as usual, has denied any similarity between his own and the Quiggin-Yaari approach. Cf. Allais (1988a: 244).

[110] Decision theorists have turned to the systematic evaluation of alternative models against empirical data only recently (see for instance the works cited in the next footnote below). The results of these exercises being rather controversial, theory appraisal is based on mixed considerations based on normative, formal and empirical criteria. It must also be stressed that decision theorists outside economics tend to see things in different ways.

[111] A similar thesis is suggested by Chris Starmer in a forthcoming survey of generalised expected utility models. Starmer (2000) sees normative requirements as impediments to the development of an empirically successful theory of decision, with which I agree; except that - by narrowing the range of possible solutions – normative principles have also paradoxically helped in the derivations of specific 'superseding' models. Whether the latter are 'successful' from a descriptive point of view, is another issue raising interesting, but different, methodological questions (cf. for example Harless and Camerer 1994, Hey and Orme 1994, and the methodological discussion in Anand 1998).

model, vNM's EU theory, was formulated, empirically refuted, normatively defended, normatively falsified, and finally modified. Syntactical, empirical, and normative considerations all contributed to the identification of the principle of independence as the 'weak axiom' in EU theory. Recent developments, however, have shown the readiness of some decision theorists to dig deeper into the empirical and normative 'hard core' of EU theory. Counterexamples to more fundamental axioms such as weak ordering have been well-known since at least the early Fifties.[112] Kenneth May (1954), for instance, performed an experiment with college students, where the subjects were asked to choose among marriage partners assessed in terms of intelligence, looks, and wealth. The hypothetical partners $\{x, y, z\}$ were ranked $x > y > z$, $y > z > x$, and $z > x > y$ according to each characteristic respectively. Intransitive patterns arose when subjects were asked to choose among pairs of potential partners. In another curious experiment reported in the same paper May asked war pilots about their (hypothetical) behaviour in case they got caught in a burning plane. Again, cyclical patterns of 'desperate choices' could be observed.

May's empirical counterexamples had little effect in shaking economists' confidence in transitivity, perhaps because the author touched the issue of normativity only incidentally,[113] because the article had nothing to do with behaviour under risk and uncertainty, and because, of course, to give up a theory's 'hard core' can be more painful than to tolerate anomalies. Several experiments tried to test intransitivity during the fifteen years that followed May's article;[114] although violations were detected, none of them was related to those of EU theory. Evidence of intransitive behaviour did not raise much interest among economists in general, not until it was widely discussed by decision theorists, at least. Amos Tversky (1969) first published a series of experimental results pointing to violations of Weak Stochastic Transitivity - a weakened transitivity axiom reformulated so as to take the possibility of random mistakes into account. One of the most striking results of Tversky's experiments was that his subjects engaged in cyclical choices even though they held the general view that transitivity is a fundamental tenet of rational choice. Tversky, a hard normativist himself

[112]Cf. the references in Fishburn (1991).

[113]May commented critically on the 'postulationist' standpoint: "Of course, the whole issue [of the descriptive adequacy of the theory] can be avoided by simply asserting transitivity as part of the definition of 'rational behaviour'. The question then is whether rational behaviour as so defined has very much importance, either descriptive or normative" (1954, p. 8).

[114]Tversky (1969, p. 32) cites at least seven papers reporting experimental results of intransitivity tests.

concerning EU theory, tried to explain the evidence by pointing to the fact (among others) that in dynamic choices sometimes agents are not committed to the original set of preferences: the elimination of an alternative can cause a change in agents' tastes (1969, p. 31).

Other anomalies and counterexamples to the principle of transitivity emerged in the following years. One of them, the phenomenon of preference reversals[115] - 'discovered' in 1971 by two psychologists, Sarah Lichtenstein and Paul Slovic - achieved what Tversky had not: it became the first instance of intransitive choices to be systematically discussed by economists. Lichtenstein and Slovic predicted, and in fact discovered, that the majority of people when facing a choice between a gamble with a high chance of winning a small sum, and another gamble with a low chance of winning a large prize would choose the former; but the same subjects, when asked to price these gambles, would sell the low bet at a higher price than the high one.[116]

Economists' first reaction was to make the counterexample disappear and 'save' the received theory. Grether and Plott (1979) tried unsuccessfully to prove the phenomenon to be an artefact of Lichtenstein and Slovic's experimental procedures. Holt (1986), Karni and Safra (1987), and Segal (1988) devised ingenious attempts to explain the anomaly away by giving up the independence axiom of EU theory (a move by then made respectable by Machina's and others' generalised models) and showing how the illusion of reversals could result from the use of certain mechanisms for eliciting preferences customarily used in these experiments. All such attempts are now considered to have failed, and it is now generally accepted that preference reversals do really occur in laboratory settings. These experiments will be illustrated and discussed in detail in the next chapter (see also the references therein). From the point of view of the present chapter, the interest of the preference reversals phenomenon lies in its having stimulated the development of generalised models of decisions without the transitivity axiom. To put it differently, preference reversals have 'hit' the hard core of both the EU model and some generalised EU theories such as those inspired by Allais' ideas.

---

[115]For an introduction to this anomaly, see Thaler and Tversky (1990).

[116]It should be stressed that this is just *one* possible reading (although a most plausible one) of the phenomenon (see next chapter for a full discussion). The discovery of preference reversals confirms that competition between rival theories is the best incentive for the growth of scientific knowledge. Lichtenstein and Slovic could predict the phenomenon because they had observed a high correlation between prices and payoffs and between gambles and chances of winning. The experiment described in their (1971) was developed in order to confirm their theory, and was only later taken seriously by economists (especially after Grether and Plott's (1979) attempt to 'explain the anomaly away').

In 1982 Graham Loomes and Robert Sugden tried to take seriously the possibility of violations of the transitivity axiom, and proposed a new theory, called 'regret theory', which could account for many counterexamples to EU axioms, including the Allais paradox and preference reversals.[117] Loomes and Sugden replaced vNM's equation (1) with the hypothesis that agents choose so as to maximise the *expected modified utility* E_i^k, defined as

$$(2)\ E_i^k = \sum_{j=1}^{n} p_j m_{ij}^k$$

Here j is an index of the actual state of the world, i an index of the actual choice made, and k an index of a possible choice which was never made. The crucial difference compared to EU theory lies in the *modified* utility function m_{ij}^k, which Loomes and Sugden define as a function of the ‘choiceless utilities’ c_{ij} and c_{kj}:

$$m_{ij}^k = M(c_{ij}, c_{kj}).$$

The idea of 'choiceless utility' apparently sets regret theory in the tradition of classical (i.e., Benthamite) utility. Loomes and Sugden think that we can meaningfully speak of utilities independently of any choice between gambles, or in other words assume the existence of utility as a psychological experience which we know about by introspection.[118] 'Choiceless' utility "is the utility that the individual would derive from the consequence [or state of the world] x if he experienced it *without having chosen it*" (Loomes and Sugden, 1982, p. 807). Intuitively, regret theory claims that in terms of pleasure/pain it makes a big difference whether a consequence is the result of a process which at some stage involved a decision by the agent or not. In the latter case we can speak of 'choiceless' utility, in the former the situation is more complicated because the agent may experience the opposite feelings of regret and rejoice according to 'what might have been'. Thus, it is important not only to take into account the utility assigned to the consequence x_{ij} of the actual choice a_i in the actual state of

---

[117]See Loomes and Sugden (1982), (1983), (1984) and Loomes, Starmer and Sugden (1991).
[118] Sugden (1993) has later devised a presentation of regret theory from a set of preference axioms. I stick in the main text to the original version of the theory because it sheds more light on the heuristics that generated it, as well as on the normative arguments that may support it.

the world s_j (i.e. c_{ij}), but also the utility of the consequence x_{kj} of the possible choice a_k in the actual state of the world s_j (i.e. c_{kj}).

If we simplify the function $M(.)$ and define a 'regret-rejoice function' $R(.)$ assigning a real-valued index to every possible increment-decrement of choiceless utility, we can easily show how regret theory takes into account EU theory as a special case.[119] It is worth noting that the heuristic path to regret theory is similar in some respects to that of Machina's generalised expected utility model. In relation to the model to be superseded, the strategy leading to the superseding theory is characterised by two steps: one of generalisation and one of restriction ('zig-zag' heuristic process). Regret theory is apparently a weaker generalisation of EU theory, because it has two free parameters instead of one: a choiceless utility function $C(.)$ and a function $Q(.)$, defined such that $Q(\xi) = \xi + R(\xi) - R(\xi)$ for all ξ. The determination of the $C(.)$ and $Q(.)$ functions (second step towards a model with richer content) is entirely an empirical matter in regret theory. It is remarkable, however, that, merely by assuming that $Q(.)$ is convex for all $\xi > 0$ and that $C(.)$ is monotonically increasing, anomalies such as the Allais paradox and preference reversals (as well as others) can be accounted for.

Given that the existence of rival theories of rational choice seems to play a crucial role in the procedure of acceptance of a counterexample to a theory of decision, it is interesting to investigate whether some theory of rationality can help to legitimise preference reversals as normative (rather than mere descriptive) counterexamples, and perhaps also legitimise regret theory as a model of rational decision making. Intransitive patterns of choices are inadmissible in traditional decision models like EU theory. Such models assume that only the consequences of an action really matter, not the comparison between them and the consequences of alternative actions under the same state(s) of the world. It is taken for granted that agents approach the choice-situation with a definite master-plan of preferences which makes each possible choice consistent with any other.[120] But we can think of situations where no such master-plan exists, and choices are highly dependent on the set of alternatives, so that a choice between $\{a_1, a_2\}$ is not necessarily consistent with a choice between $\{a_2, a_3\}$ and a choice between $\{a_1, a_3\}$. This is the particular case regret theory is designed to explain. The relation $>_R$ in regret theory is not equivalent to the relation $>_{EU}$ in EU

---

[119] Assuming that $m_{ij}^k = c_{ij} + R(c_{ij}, c_{kj})$, if $R(\xi) = 0$, then $m_{ij}^k = c_{ij}$. Cf. Loomes and Sugden (1982, p. 809).

[120] See Sugden (1985).

theory: $a_1 >_R a_2$ does not mean 'having a_1 is preferred to having a_2', but rather 'choosing a_1 and rejecting a_2 is preferred to choosing a_2 and rejecting a_1'.[121]

Normativists supporting EU theory and generalised EU analysis share the view that some restrictions must be imposed on preferences in order to make them 'rational': among these restrictions, there are transitivity and completeness. But are such restrictions necessary from a normative point of view?[122] Can we not claim that even Allais' ‘minimal’ version of the rationality principle is just a specific formulation of a more general underlying principle? Loomes and Sugden's theory may pass the Savage-Allais quasi-empirical test of rationality, if subjects acknowledged that rationality is a feature of choice once beliefs and desires are given*:* "a choice may be rational or irrational, but an experience [e.g. regret or rejoice] is just an experience" (Loomes and Sugden, 1982, p. 820).

> In what way are elation and disappointment less rational sources of pleasure and displeasure than, say, winning a race or having a manuscript rejected by a publisher? Economists have not normally thought it their job to classify some sources of utility as 'rational' and others as 'irrational'. Our belief is that sensations of pleasure and pain are simply *sensations*, and that the issue of rationality is not about what kinds of sensations are felt, but about how the individual responds to them (Loomes and Sugden, 1984, p. 227).

In the light of regret theory, even Allais' 'minimal' principle of rationality turns out to be not so 'minimal', after all. The 'ultra-minimal' principle to which Loomes and Sugden appeal[123] is anchored in an old and prestigious tradition, running from David Hume (1740) to Karl Popper (1967). According to that tradition,

> Reason is, and ought only to be the slave of passions (Hume, 1740, p. 415).

Neo-Humeans retain point (b) of Allais' definition of rationality[124] - that agents employ means that are appropriate for their ends - and reject (a), i.e. do not impose any normative restriction on individual desires. According to this interpretation, such restrictions can at most be taken as fallible empirical

121 See Loomes and Sugden (1982), p. 821.

122 For an extended discussion of this issue, cf. Anand (1993, ch. 4).

123 See also Sugden (1991).

124 See above, section 2.3.

hypotheses on early conjectural models, to be later modified in the light of new evidence about behaviour deviating from our models.

2.6. Conclusions

‘Normativism’ and ‘descriptivism’ have been strictly entangled in the history of decision theory, and this has had some important consequences. First of all, it is impossible to give an account of the development of EU theory as an empirical theory of human behaviour without taking the problem of rationality (in all its subtle aspects) into account. If on the contrary we assume that the problem of normative justification was primary in the controversy then we can restore a fuller picture of the reasons that guided changes in the *content* of the theories of decision that replaced the received model during the last twenty years. Here, I think, the fruitfulness of Lakatos' (1963-64/1976) framework can be best appreciated for its ability to explain the dialectics of counterexamples, refutations and redefinitions of the formal theory at stake. Heuristics in the mathematical and in the *normative* sciences seem to be founded on a similar mechanism: the formal developments of the discipline are in the forefront, but ‘informal’ theories guide them from behind the scenes. I have shown that shifts in the concept of rationality anticipated or immediately followed empirical refutations of positive models of decision, thus providing a set of standards to be fulfilled by superseding theories, and a normative justification to the anomalous evidence produced in the experiments. The debate on normativism can be reconstructed as a struggle for the legitimacy of counterexamples and alternative models of rational behaviour.

Earlier reconstructions of the development of decision theory have been concerned either with formal aspects only (Fishburn and Wakker, 1995), or have projected onto the subject matter a framework of analysis taken from general philosophy of science (Mongin, 1988a). It turns out that the axioms of the received model to be protected from refutation are the same whether identified by a normative argument grounded in rational choice theory (as stressed in my approach) or by a descriptive (i.e., growth-of-knowledge) argument (as in Mongin, 1988a). The two sets of arguments by and large account for the theoretical changes that actually took place. The difference between the two reconstructions lies in the *content* of the heuristics followed by anti-EU theorists, rather than in its *consequences*.

The idea that normative models of behaviour play a heuristic role in the social sciences is of course not new, but has usually been stated in rather imprecise ways, by saying that rational choice models are ‘ideal-types’, or false but useful ‘first approximations’ to real human behaviour. In this chapter, I have tried to clarify the status of the rationality principle and of its role in improving on earlier imperfect models of choice. Most importantly, I have suggested that, in the theory of decision under risk, *different* rationality principles have shaped the standards of acceptance of theories in different periods, have implicitly dictated which components to revise in the light of counterexamples, and how. A ‘logic of normative falsification and theory-improvement’ has therefore provided decision theorists with a rather rigorous, quasi-deductive procedure for scientific discovery.

So far I have looked at rather simple experiments (whose results, nonetheless, proved to be remarkably robust). In the years that followed Allais’ early tests, experimentalists have developed sophisticated laboratory techniques in order to extend the range of their investigations. Such techniques, however, have also made the interpretation of the experimental results more complicated. Often, the inference from observed data to a phenomenon of interest involves highly theoretical assumptions that are as tricky as the hypotheses under test. In the next chapter I shall examine a case of the latter sort, in order to show how problems of interpretation can be solved by means of more experiments. I shall suggest that the in-principle possibility of further testing in controlled conditions is one of the strengths of experimental economics, and the main reason why it should be preferred to empirical testing ‘in the field’.

Chapter 3

Phenomena and Artefacts
Preference reversals and the Becker-DeGroot-Marschak mechanism

> 'The art of representation is therefore a long way removed from truth, and it is able to represent everything because it has little grasp of anything, and that little is of a mere phenomenal appearance. For example, a painter can paint a portrait of a shoemaker or a carpenter or any other craftsman without understanding any of their crafts; yet, if he is skilful enough, his portrait of a carpenter may, at a distance, deceive children or simple people into thinking it is a real carpenter.'
>
> (Plato, *The Republic*, Part 10, 598b)

3.1. Introduction

Controversies in economics often fizzle out unresolved. One reason is that, despite their professed empiricism, economists find it hard to agree on the interpretation of the relevant empirical evidence.[125] In this chapter I will present an example of a controversial issue first raised and then solved by recourse to laboratory experimentation. A major theme of this chapter, then, concerns the methodological advantages of controlled experiments. Experimentalists

---

[125] I am not suggesting that this is the *only* reason; for more views on this topic, see the symposium in the *Journal of Economic Methodology*, Vol. 1, June 1994 (with articles by Thomas Mayer, Deirdre McCloskey, Roger Backhouse, David Colander and Henry Woo).

themselves have put forward some justifications for their laboratory turn. John Hey (1991), for instance, focusing on the traditional theory-testing role of experimentation, argues that any econometric test is a test of two components: a theory (which is supposed to involve a *ceteris paribus* clause), and a set of assumptions about the structure of the 'disturbing' factors not modelled in the theory. Thus,

> if 'the theory' survives the test, it could be because *both* the original economic theory *and* the assumptions about the stochastic variables are correct, or because *both* the original economic theory *and* the assumptions are incorrect. There is no way of telling which. Similarly, if 'the theory' does not survive the test, there is no way of telling whether this is because the economic theory is correct and the stochastic assumptions incorrect, or because the economic theory is incorrect and the stochastic assumptions correct, or because both are incorrect. Hence, a conventional econometric test of some economic theory is not really a test of that theory at all (1991, p. 8).

Such an argument will sound very familiar to philosophers of science: it starts from a version of the so-called Duhem-Quine problem. Hey goes on to suggest that controlled experiments may help to solve the problem. Hey is pointing in the right direction, but theory-testing is just *one* aspect of experimenting, and experimenters are often merely concerned about determining whether a certain phenomeonon exists or not, or whether, when, and where it can be produced, without necessarily engaging in any theoretical explanation of the phenomenon itself. In this chapter I shall be concerned mainly with such a case, and focus on the example of preference reversals, a phenomenon whose existence was until quite recently denied by the majority of economists. Their favourite strategy consisted in trying to explain the phenomenon away as an artefact of the experimental techniques used to observe it. By controlled experimentation, as we shall see, such an interpretation has been discredited, and now preference reversals are generally accepted as real.

The second theme of this chapter is then the nature of experimental artefacts and of the experimental methods devised to detect them. The problem of distinguishing an artefact from a real phenomenon will be shown to be related to methodological issues traditionally discussed by philosophers of science, such as the theory-ladenness of observation and the Duhem-Quine problem. A large part of this chapter is devoted to clarifying these two philosophical problems, and to arguing that only the latter is relevant to the case at hands. The solutions devised by economic experimentalists to the Duhem-Quine problem will be presented and discussed. I shall show that they belong in two broad categories: independent tests of new predictions derived from the competing hypotheses at stake, and 'no-miracle arguments' from different experimental techniques delivering converging results despite their being theoretically independent. I shall not try to defend the rationality of such procedures and arguments, but rather aim to give 'a concrete account of scientific life', thus accepting Ian Hacking's challenge in his *Representing and Intervening*:

> If we want a concrete account of scientific life we should consider what, for example, it is to have an experiment working well enough that the skilfull experimenter knows that the data it provides may have some significance. What is it that makes an experiment convincing? (Hacking, 1983, p. 181).

3.2. Experiments

There are two main and related reasons for experimenting in the laboratory: replicability and control. *Replicability* fulfills one of the basic standards of science: the requirement of publicity. Whenever a scientific result has been asserted, it should be possible for different researchers to check the result by replicating it in a different laboratory. This procedure aims at reducing the bias of subjective elements in scientific knowledge. The second advantage of laboratory experimentation is that, by making it possible to keep a system *fixed* in all respects deemed relevant by general considerations, it permits the acquisition of large amounts of data about 'the same system' (whereas in the 'wild', of course, there may be important uncontrolled variations so that successive observations are *not* of the same system). Suppose, for example, we are interested in testing the adequacy of a theoretical model to explain some puzzling empirical evidence. The model will try to articulate at least partially the processes generating the data to be explained; the pattern of data will be derived from the model when particular values are assigned to its free parameters and variables. A natural way to test the validity of the model is to check whether the observable features of the system change in accordance with the predictions of the model when diffferent initial conditions are assigned to the variables (keeping the parameters constant). We need, in other words, different sets of data generated by the same system under different initial conditions. Outside the laboratory this kind of test is problematic, because there always is the possibility that two different sets of data be generated by two different 'data generating processes'. Roger Backhouse (1997) has called it 'the econometrician's regress':[126]

> To confirm that we are measuring an economic phenomenon correctly we have to show that it occurs in more than one data set drawn from the same population; to know whether our data sets are drawn from the same population we need to show whether they exhibit the same phenomenon; to establish whether they exhibit the same phenomenon we have to be able to measure it correctly, and so on (Backhouse, 1997, pp. 146-147).

In the laboratory, one can usually *control* for the initial conditions more tightly, so as to (i) make sure that the data generating process is held fixed, and (ii) intervene to change the values of the main theoretical variables and thus test the robustness of the relationships at stake. Replicability and control provide the means of interpreting empirical evidence in a more straightforward way than field research does - by discriminating, for instance, between two alternative theoretical explanations of the same data.

Experimentation can therefore help to check the quality of the data. The treatment of data is a crucial problem in econometrics. Econometricians have often access only to 'bad' data (and have a limited number of them at that), and there is always the danger of performing incorrect pre-filtering of data, stochastic misspecification, errors in aggregation, or of having chosen inadequate sample-sizes. The correct prediction or explanation of phenomena is therefore dependent on two main factors: a correct theory of the phenomenon at stake, and a correct interpretation and treatment of the data available. Although the quality of the data collected in the laboratory is usually much better than that of field data, the same problem holds to different degrees in both cases. Laboratory data can be misinterpreted when the experimental procedures used to obtain them are not understood. This is *the* problem of 'artefacts', and I shall discuss it in detail in the following sections suggesting that the techniques experimental economists use to solve it are representative of the strategies used in general to support experimental claims in science. Artefacts are due to mistaken inferences from data, as I shall try to show below. Since data are more messy in field 'experiments', it is easier to produce artefacts there than in the laboratory. In this chapter, a number of techniques to detect artefacts will be illustrated by means of concrete examples of experimentation and then discussed. The scope is to let the techniques which

[126] After Harry Collins' (1985) 'experimenter's regress'.

render laboratory experimentation attractive emerge from a description of the practice of science.

The case of preference reversals discussed below is not new to philosophers of science: Hausman (1991; 1992) has used it as an example of economists' dogmatic attitude towards disturbing empirical results. Tammi (1997) has analysed the debate from an argumentative point of view in a way designed to uncover economists' tacit commitments, presuppositions, and rhetoric. Here I shall focus mainly on the problem of interpreting experimental results when inferences have to be drawn to a phenomenon of interest from data that are produced by a rather complicated apparatus.

3.3. Preference reversals

Neoclassical economic theories describe the properties of preference scales. A number of techniques for eliciting preferences have been developed by experimentalists, mostly based on the notions of willingness to pay. According to standard economic theory, the results of these operations should all be consistent, the behaviour of economic agents being in all these measurement contexts determined by their preferences. Some experimental psychologists began in the late sixties to question the very existence of preference scales. They conjectured that, far from constituting the stable substratum from which all economic behaviour arises, preferences display a much more unstable structure and depend heavily on the situation: they are 'constructed' and vary from context to context.

Paul Slovic and Sarah Lichtenstein, two psychologists at the Oregon Research Institute, designed a two-stage test, later to become famous as the 'preference reversal experiment'.[127] I have already introduced these experiments in the last chapter, but it is worth recalling their main features. Agents were asked in separate tasks to choose among two bets and to price them. The pairs of lotteries had a common feature: they consisted of a bet with a high probability of winning a moderate amount of money and a low probability of losing a small amount (called the 'P-bet'); and a bet with a low probability of winning a larger sum and a high probability to lose a smaller sum (the '$-bet').[128] Moreover, they had approximately the same expected monetary value. Slovic and Lichtenstein's conjecture was that "bidding and choice involve two quite different processes that

[127] Lichtenstein and Slovic (1971), (1973).

[128] For instance: P-bet = (.9 to win $5, .1 to lose $1); $-bet = (.3 to win $15 and .7 to lose $2).

involve more than just underlying utilities of the gambles" (Lichtenstein and Slovic, 1971, p. 47).

In previous studies, Slovic and Lichtenstein (1968) had observed a high correlation between, on the one hand, prices and payoffs, and, on the other, choices and probabilities. An experiment was conceived explicitly to produce patterns of choices such that the agents chose the P-bet but bid more for the $-bet. As a matter of fact, such patterns were observed, and have since then been known as instances of the 'preference reversal phenomenon'. The standard rate of reversals observed by Lichtenstein and Slovic, and then in later preference reversal experiments, was between 70 and 80%. Not all reversals were of the kind predicted by Lichtenstein and Slovic, though: in a 'standard' PR experiment, from 15 to 25% of reversals are of the non-predicted (or 'asymmetric') type.

Lichtenstein and Slovic (1971) performed three experiments. In order to control for possible disturbances due to lack of incentives, they used in two of their experiments an elicitation procedure known since the mid-sixties as the Becker-DeGroot-Marschak (BDM) mechanism. The BDM procedure is a tool devised to elicit the selling price of any kind of commodity, and as such has been often used to control subjects' preferences over *lotteries*. To elicit the certainty-equivalent of a lottery, in fact, a pay-off mechanism must be used to make sure that the price reflects the subject's *real* preference.

In a BDM elicitation, a subject is asked to state her reservation price, s, for a lottery (say, $[x, p; y, (1 - p)]$); then, the lottery is auctioned, and if a buyer willing to bid a sum $b \geq s$ is found, the subject receives b; otherwise, the lottery is played, and the subject receives a sum x or y according to the outcome. The experimenters draw the bidding sum b from a uniform distribution over some relevant set. It is easy to show that a rational utility maximiser must state his real selling price.[129]

The BDM mechanism is often used in conjunction with the so-called Random Lottery Selection (RLS) procedure. In general, experimental subjects are asked to perform a number of tasks; instead of receiving an aggregate payment, the subject is rewarded according to the results of only *one* task selected at random. This procedure controls for endowment effects (when a subject is asked to perform several tasks, her preferences may vary because of changes in her

[129] Cf. Becker, DeGroot and Marschak (1964).

wealth) and reduces experimental costs at the same time. If the selected task is a choice one, it is simply played out; if it is a pricing task, the BDM mechanism is used.

When other experimentalists began to replicate Lichtenstein and Slovic's findings, they also used the BDM and RLS procedures. David Grether and Charles Plott, the first economists to take PRs seriously, investigated the replicability of the PR phenomenon at the California Institute of Technology in the late seventies. Their research was driven by the suspicion that PRs may have been the product of some undetected experimental effect. In their (1979) paper Grether and Plott list thirteen possible sources of 'disturbance', pointing therefore to thirteen possible ways to account for (or 'explain away') PR data. These included misspecified incentives, income effects, indifference, strategic responses, ill-defined subjective probabilities, elimination by aspect, lexicographic semiorder, costs of information processing, confusion and misunderstanding, unsystematic and sporadic behaviour, unsophisticated subjects, experimenter's effects. Despite great care in designing the experiment to control for such disturbances, the experimenters observed the same results Lichtenstein and Slovic had produced a few years earlier. The historical-methodological significance of Grether and Plott's experiment has been discussed in depth by Dan Hausman (1991; 1992), and I shall refer to his work without further comments.[130]

3.4. Data and Phenomena

Bogen and Woodward (1988) have forcefully argued that scientists customarily explain *phenomena*, rather than data. Phenomena can be thought of as similar to what neopositivists called 'experimental laws',[131] regularities occurring in some specific experimental situation. Like many experimental laws, Bogen and Woodward's phenomena are not directly observable. They are rather *inferred from data*.

> Data, which play the role of evidence for the existence of phenomena, for the most part can be straightforwardly observed. However, data typically

---

[130] For a non-technical presentation of the early research on the PR phenomenon, cf. also Tversky and Thaler (1990).

[131] Cf. Nagel (1961), ch. 5.

> cannot be predicted or systematically explained by theory. By contrast, well-developed scientific theories do predict and explain facts about phenomena. Phenomena are detected through the use of data, but in most cases are not observable in any interesting sense of the term (Bogen and Woodward, 1988, pp. 305-306).

The preference reversals phenomenon is an example of a 'phenomenon' in Bogen and Woodward's sense. To begin with, PR are not directly observable. We can only observe patterns of behaviour that appear *prima facie* incompatible with the claim that 'there exists a transitive scale of preferences underlying subjects' choices'. The data obtained in a typical 'preference reversals experiment', in contrast, may be represented as sentences like 'subject *x* has chosen the P-bet over the $-bet and priced the $-bet higher than the P-bet'. In order to obtain the PR phenomenon, one needs to assume, to begin with, that pricing and choosing convey genuine information about preferences.

At this stage psychologists and economists part company. Economists, in fact, also presuppose that the same preference structure underlies both pricing and choosing, whereas psychologists - as I have already mentioned - doubt that the idea of a stable preference scale is useful at all.[132] Following the 'economic' approach one is led to infer the existence of a genuinely intransitive preference structure. The phenomenon can then be represented as follows:

(PR) $P >_c \$ >_p P$, and $>_c = >_p$.

Some theorising has thus taken place on the way from the observation reports to the phenomenon in the form above. '$>_c$' stands for 'preference as emerging from choice', and similarly '$>_p$' for 'preference emerging from pricing'. The first inference involves some assumptions about the correct functioning of our instruments of elicitation, whereas the latter involves a commitment to the principle of *procedure invariance* - the idea that all economically relevant behaviour is determined by the same preference scale, and thus that all economic behaviour can be used as evidence for inferring the structure of preferences. Theoretical and

[132]For reasons of simplicity, in this paper I shall mainly focus on the 'economic' interpretations of the PR phenomenon, thus disregarding those which question the very existence of a preference structure. I shall therefore often identify the PR phenomenon with intrasitive preferences rather than (more correctly) with the non-existence of a transitive preference scale in general. For some attempts to discriminate in the laboratory between the 'economic' and the 'psychological' interpretations, see Loomes, Starmer and Sugden (1989) and Tversky, Slovic and Kahneman (1990).

non-theoretical assumptions of this kind sanction the step from reports like 'subject *x* has chosen so-and-so' while 'subject *y* has priced so-and-so' to claims about preferences; or from observed apparent 'price reversals' to real preference reversals.

It is interesting to notice that the term 'preference' is used by both economists and psychologists when debating the results of PR experiments. Ironically, the 'preference reversal' label was invented by psychologists, despite the fact that they do *not* believe in preferences. Psychologists use the term in its 'commonsense' meaning, and originally would not distinguish between 'price-choice reversals' (the data) and 'preference reversals' (the phenomenon). Nevertheless, they did not oppose the economists' shift to a slightly more technical connotation. To economists, PR are data seen through the filter of what we may call the 'beginning' of an explanation – the low-level assumption that it makes sense to speak of preferences in the first place. This presuppostion is still far removed from a full theoretical explanation of the data, an explanation that in the PR case would involve some precise claim about the *structure* of preferences.

The PR case can be used as a counterexample to a popular interpretation of Kuhn's *Structure of Scientific Revolutions* (1962/1970), according to which scientific terms and concepts take completely different meanings depending on the 'paradigm' they belong to. This thesis is behind the famous Kuhnian claims that different paradigms are 'incommensurable' and scientists from different traditions 'live in different worlds'. We shall soon discuss at length other aspects of this view. For the time being, let us just keep in mind that the experimental activity we are going to review in the following sections illustrates how communication across paradigms *can* take place, and be profitable too. There exists a 'trading zone', to say it with Peter Galison (1997), where scientists negotiate meanings that allow communication and often the resolution of disputes.

3.5. Artefacts

The distinction among data and phenomena should help us introducing some further conceptual artillery to be used in the next sections. Let us start from the concept of '*artefact*'. The word 'artificial' has an ambiguous meaning: it stands in opposition to 'natural', a word with clear normative flavour, and for this reason has a definitely negative connotation. An artificial object may be an imitation of an original, and as such not genuine, not 'true'. The artificial can be 'deceitful', 'insincere'. In science the term is used in a number of ways, but I shall here be concerned with its use in relation to phenomena only. An interpretation is an artefact when it is not true, a mere illusion of the instruments of observation. Artefacts are a case of potentially misleading connection between data and phenomena.

Microscopy textbooks for example report a large taxonomy of artefacts which the student is likely to encounter in laboratory work. Bubbles on a slide, fringes caused by optical aberrations around the edges of a cell, stains, scratches, folds produced during the preparation of the assay and so on. For instance, if the membranal border of an organelle seems to be interrupted somewhere, this may be due to the chemicals used to preserve the tissue. The 'natural' membrane was continuous, but the chemical substances used by the experimenter caused its deterioration.[133] Not being aware of this fact, the experimenter might infer that it was a characteristic of the cell ahead of any intervention on her part.

Artefacts are negative features of data for a number of reasons. They sometimes render the analysis of the phenomenon at stake very difficult if not impossible by irreversibly blurring it or covering it. But above all, artefacts may be taken as 'natural' properties of the investigated entity. As such, the important notion of 'artefact' is intrinsically epistemic: an artefact results from our ignorance of the effects of the experimental procedure, which we have not been able to control properly. If the concept of artefact is an epistemic one, then there must be a step where we miss something. In particular, *an artefact is a feature produced by the method of observation that may lead to incorrect interpretations of the data*. For instance, when the data are contaminated by some unknown factors; or when one has an incorrect theory about the functioning of an instrument. In the example above, the artefactual phenomenon is the break in the membrane of a certain organelle, and the incorrect inference that led to it was suggested by the visual display of the gap in the membrane (the data), and by our ignorance of the chemicals' side-effects. We shall see that economists talked about the 'artificiality' of the preference reversals phenomenon in a similar way, at least in the first stage of the controversy.

3.6. Explaining preference reversals away[134]

[133] Cf. Lynch (1985), ch. 4, for a number of examples from neurobiology.

[134] Although I have tried to simplify the issue as much as I can, the contents of this section remain quite technical; it should be possible however to follow the general line of argument while skipping some of the details.

In the mid-eighties some economists began to argue that the PR phenomenon could have been an artefact of the experimental procedures used in order to control subjects' preferences: in other words, that inferences from data to $P >_p \$$ and $\$ >_c P$ were faulty. Charles Holt on the one hand, and Edi Karni and Zvi Safra on the other, independently and almost simultaneously, began to investigate theoretically the robustness of experimental procedures to violations of the axioms of expected utility theory. These authors pointed out that the controls used by Grether and Plott and other experimenters[135] "are appropriate if the axioms of von Neumann-Morgenstern utility theory are satisfied" (Holt, 1986, p. 509). Their argument were directed towards either the BDM procedure or the Random-Lottery Selection (RLS) procedure. The dependence of elicitation procedures on expected utility theory was hardly a new discovery: the inventors of the BDM mechanism knew and wrote explicitly that "the procedure is based upon the [...] well-known 'expected utility hypothesis'" (Becker, DeGroot and Marschak, 1964, p. 226).

As we have seen above, in a PR experiment the subjects are asked to perform a number of tasks; from these, one is selected at random (RLS procedure). If the chosen task is a choice one, it is simply played out; if it is a pricing task, the BDM mechanism is used. According to the probability calculus, a lottery in multiple stages can be reduced to a single-stage one, and expected utility theory requires that people's preferences in the multi-stage lottery are consistent with those in the reduced one. Formally, the reduction principle states that subjects are indifferent between a compound lottery $A = (X_1, q_1; \ldots; X_m, q_m)$, giving a chance q_i to participate in a lottery $X_i = (x_1^i, p_1^i; \ldots; x_{ni}^i, p_{ni}^i)$, and the reduced lottery $R(A) = (x_1^1, q_1 p_1^1; \ldots; x_{n1}^1, q_1 p_{n1}^1; \ldots; x_1^m, q_m p_1^m; \ldots; x_{nm}^m, q_m p_{nm}^m)$.

[135] Cf. Pommerehne, Schneider and Zweifel (1982), and Reilly (1982).

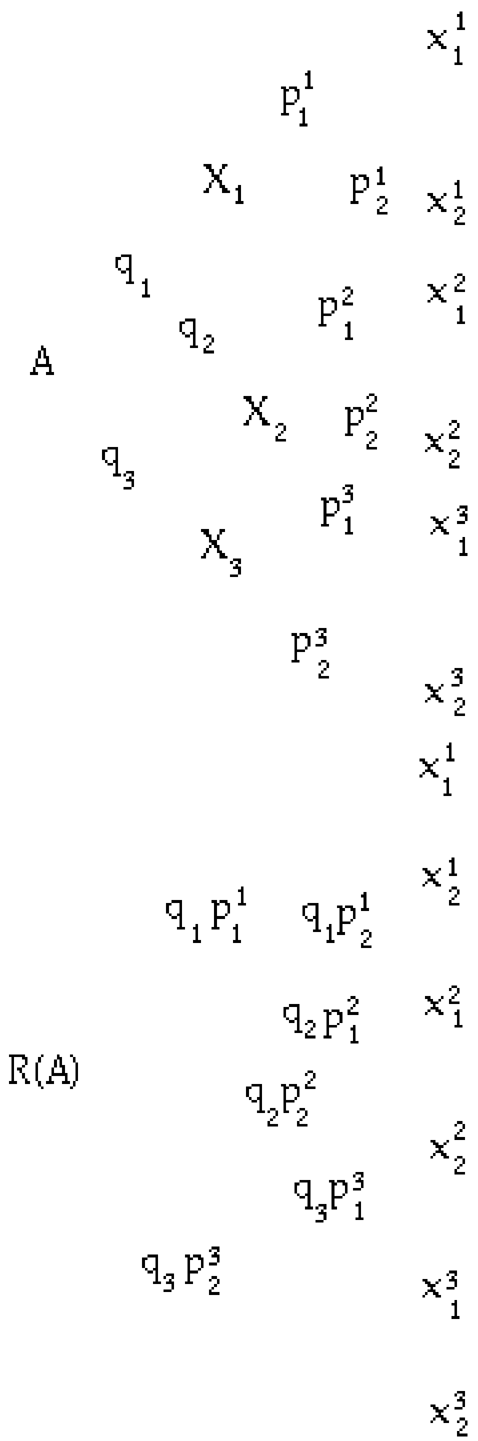

Figure 1: A compound lottery A and its reduced counterpart R(A)

Holt conjectured that subjects saw the preference reversal experiment as a two-stage lottery: in the first stage, the task to be played out is randomly selected; in the event of pricing, there is a second stage, i.e. the task is played out *via* the BDM mechanism. Then, Holt showed that *if* they apply reduction but not the independence principle, some subjects who *prefer* the \$-bet to the P-bet may reverse their *choices* during the experiment. Choices observed *via* RLS mechanism, then, may not reveal true preferences.

The suggestion of blaming violations of independence was in line with the arguments put forward in that period by theorists working on other anomalies. Mark Machina's (1982) 'Generalised Expected Utility Analysis' (GEUA) was at the time a fully developed alternative approach to decisions under risk, which relaxed the principle of independence and allowed utility functions to be merely differentiable rather than strictly linear in the probabilities. Other approaches, like Chew and MacCrimmon's (1979) 'Alpha Utility' theory, or Quiggin's (1982) and Yaari's (1987) 'Expected Utility theory with Rank-Dependent Probabilities', were being developed which similarly made do without independence or related principles (see chapter two).

Loosely, the principle of independence says that only the outcomes that distinguish two lotteries are relevant to the decision to be taken. More precisely, it says that if a lottery X is preferred to another lottery Y, then the compound lottery (X, p; Z, 1-p) is preferred to (Y, p; Z, 1-p). As I have shown in the previous chapter, violations of this principle were the first to be discussed by decision theorists, thanks mainly to Maurice Allais' (1953/1979) early experiments. It was therefore possible to argue that the illusion of preference reversals resulted from violations of the Allais kind.[136] According to Karni and Safra (1987), in particular, the BDM mechanism may be perceived by subjects as a two-stage lottery giving, among its outcomes, the possibility of playing out the priced gamble. Suppose the latter is X = (4, 35/36; -1, 1/36) - one of the P-bets used by Grether and Plott (1979), typically subject to preference reversals. If, by assumption, both the $\pi(X)$ - the real selling price - and b - the bidding price - are restricted to the 1000 different values 0, 1/100, ..., k/100, ..., 9.99 ($0 \leq k <$ 1000), the following two-stage lottery results from the BDM procedure:

[136] I follow here the presentation given by Keller, Segal and Wang (1993).

$$\mathrm{A} = \left(\left(4,\frac{35}{36};-1,\frac{1}{36}\right),\frac{\pi(\mathrm{X})}{10};\delta_{\pi(\mathrm{X})},\frac{1}{1000};\delta_{\pi(\mathrm{X})+0.01},\frac{1}{1000};\ldots;\delta_{9.99},\frac{1}{1000}\right)$$

where the δ_i stand for degenerate lotteries with probability 1 of getting *i*, and $\pi(X)/10$ is the probability of participating in *X* according to the BDM mechanism. The lottery *A* is equivalent to the tree in figure 2.

By definition of a certainty equivalent (*CE*), we know that X ~ $\delta_{CE(X)}$. Thus, by applying independence, there follows that

$$A \sim A' = \left(CE\left(4,\frac{35}{36};-1\frac{1}{36}\right),\frac{\pi(X)}{10};\delta_{\pi(X)},\frac{1}{1000};\delta_{\pi(X)+0.01},\frac{1}{1000};\ldots;\delta_{9.99},\frac{1}{1000}\right)$$

The indifference above implies that agents see Tree 1 as equivalent to the tree in figure 3.

The task faced by an agent participating in a BDM experiment, then, is representable as a maximization problem: what is the value of $\pi(X)$ that maximises the value of the lottery *A'*? An expected utility maximiser, as Becker, DeGroot and Marschak (1964) had shown, will set $\pi(X) = CE(X)$. Now, by *reduction*, we can obtain

$$A \sim R(A) = \left(\left(4,\frac{35\pi(X)}{360};-1,\frac{\pi(X)}{360}\right);\delta_{\pi(X)},\frac{1}{1000};\delta_{\pi(X)+0.01},\frac{1}{1000};\ldots;\delta_{9.99},\frac{1}{1000}\right)$$

with *R(A)* corresponding to the tree in figure 4.

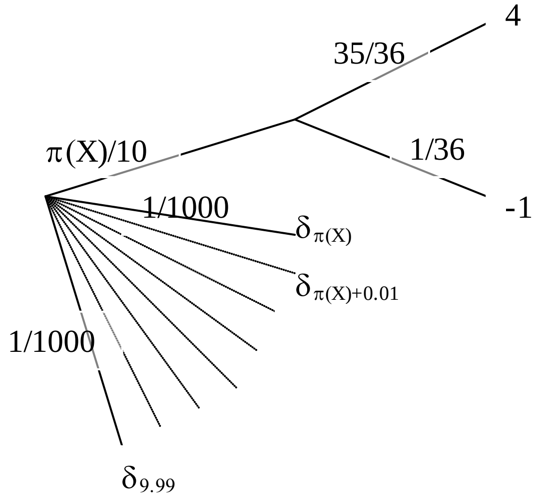

Figure 2: The lottery A

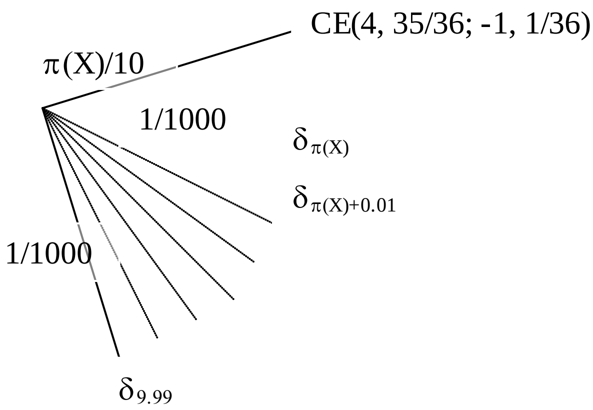

Figure 3: The lottery A’

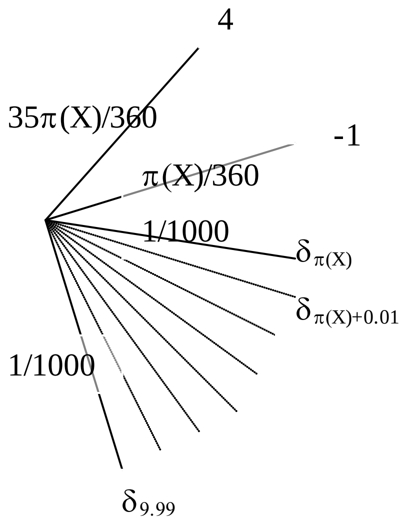

Figure 4: The lottery R(A)

Karni and Safra (1987) argued that if the independence principle is *not* obeyed, then it is not true that *always* setting $\pi(X) = CE(X)$ maximizes the value of $R(A)$. A number of generalised theories which make do without independence (Karni and Safra call them 'Ω-theories', and I have given a partial list above) can in principle be used to explain reversals. Karni and Safra went further by putting forward an example of how preference reversals were to be *expected* in the light of Quiggin's and Yaari's generalised model, 'Expected Utility with Rank-Dependent Probabilities' (EURDP), given a particular class of lotteries. The very pattern of choices observed by Lichtenstein and Slovic, Grether and Plott and others - as we shall see in section 10 below - can be accounted for by applying EURDP to the BDM elicitation. If agents were EURDP maximisers, the data produced by means of the BDM mechanism would not necessarily be inconsistent with the transitivity of the underlying preferences, and the preference reversals illusory. Karni and Safra (1987) showed also that a large class of BDM-like devices would be useless for eliciting non-linear preference relations.

By focusing on the *reduction principle,* Uzi Segal (1987) argued that violations of independence may not be the only causes of the PR 'illusion'. His argument, again, rests on the assumption that the agents perceive their task as a two-stage lottery,

$$\left(CE(X), \frac{\pi(X)}{10}; \lambda, 1-\frac{\pi(X)}{10}\right),$$

where λ is a uniform distribution on the $[\pi(X), 9.99]$ interval. Segal's second step consists in conjecturing that agents do not satisfy *reduction*, and in conceiving an example constructed on particular pairs of bets: in some cases, again, a subject may price items in a way that would not reveal her true preferences.

The general epistemic problem highlighted by the critiques of Holt, Karni and Safra, and Segal, is one of *circularity*. The 'instruments of observation' (elicitation) used in the experiments on individual choice rely heavily upon those theories of behaviour in whose investigation they are involved. Mechanisms such as the BDM procedure work by constructing further problems of choice under risk of the same kind as those under test. There is clearly a problem of circular validation here: the phenomenon at stake is inconsistent with EUT, but the instruments used to observe the phenomenon are constructed on the hypothesis that EUT is correct. Is such a circle a vicious one? And if it is, how can it be escaped?

3.7. Theories and instruments

Gaston Bachelard once claimed that "*Un instrument, dans la science moderne, est véritablement un théorème réifié*" (Bachelard, 1933, p. 140), and the case of elicitation procedures seems to instantiate this remark perfectly. The term 'reified' might even be an exaggeration: the material features of this kind of apparatus are minimal, compared to those of the instruments one finds in a physics' laboratory. There is more 'software' than 'hardware'. Experimental economics' instruments *are* theorems, perhaps.

The philosophical problem arising from a circularity between theory of the instrument and theory of the phenomena has often been discussed under the label of 'theory-ladenness of observation'. PRs constitute evidence against the existence of well-ordered preference structures, but one cannot trust what is 'seen' through the elicitation instruments because their functioning presupposes expected utility theory - a theory, moreover, for which there is established and extensive evidence of empirical violations. The physicist and philosopher Pierre Duhem once remarked that

> An experiment in physics is the precise observation of phenomena accompanied by an *interpretation* of these phenomena; this interpretation substitutes for the concrete data really gathered by observation abstract and symbolic representations which correspond to them by virtue of the theories admitted by the observer (Duhem, 1906, p. 147).

The same, as we have seen, applies to the case we are concerned with - and probably to every experimental investigation. A number of inferences lead from the data collected in the laboratory to the statement denoting the 'PR phenomenon'. Among the theories admitted by the observer there are some about the instruments used in the experiment. If the latter are challenged, the interpretational device breaks down. What are we allowed to infer from the data? Is not the very existence of the PR phenomenon challenged?

3.8. Theory ladenness

Cases of 'rediscovery' are common in philosophy. According to the mild neopositivism which shaped the standard view in philosophy of science, 'experimental laws' (the regularities produced in experimental situations - 'phenomena', in Bogen and Woodward's jargon) are independent of the truth of any particular scientific theory.[137] They constitute the neutral, solid bedrock upon which theories are constructed, and against which competing programmes are appraised. The reaction against neopositivism partly took the form of a revival of Duhem's thesis of theory-dependence of experimental phenomena. Empirical evidence is supposed to test theories, but in fact - anti-positivists began to argue - depends itself on theories. According to Karl Popper, one of the earliest anti-positivists,

> Not only the more abstract explanatory theories transcend experience, but even the most ordinary singular statements. For even singular statements are always *interpretations of 'the facts' in the light of theories*. (And the same holds even for 'the facts' of the case. They contain *universals*; and universals always entail a law-like behaviour.) (Popper, 1934/1959, Appendix *x)

How can we accept observational reports as the objective bedrock upon which theory-appraisals should be based, then? The circle seems to be vicious. Norwood Russell Hanson in his *Patterns of Discovery* (1958) introduced into the literature the expression 'theory-ladenness of observation'; Thomas Kuhn, a few years later, argued in a famous chapter of his *Structure of Scientific Revolutions* that scientists interpreting empirical data in the light of different theories 'see different things' and 'seem to work in different worlds' (1962/1970, ch. 10). I shall analyse such statements in more detail below. But *how* and *where* does theory infect observations? Theory-ladenness arguments can (and do) take different forms. For example:

a. Scientists with different theoretical presuppositions *see* different things.
b. Scientists with different theoretical presuppositions *look for* different things.
c. Scientists with different theoretical presuppositions *observe* different things.
d. Scientists with different theoretical presuppositions *report* different things.
e. Scientists with different theoretical presuppositions *infer* different conclusions from the same observations.

---

[137] See Nagel (1961), p. 87.

The taxonomy mirrors some relevant activities related to the collection and use of empirical evidence. First, experimenters usually *design* an experiment; they put a number of theoretical assumptions at work in order to produce a phenomenon of interest and relevant to their research. Sometimes scientists observe naturally occurring phenomena, but more often they must carefully arrange the experimental set-up in order to observe from a 'convenient' point of view. Then, the experimenter lets the instruments run and *sees* something: he looks at some figures on a scale, looks at some pictures taken in a bubble-chamber, or records subjects' assertions concerning how much they are willing to pay for some items. 'To see' means here 'to receive some stimuli from the environment'.[138] The experimenter may also see that a few of his subjects are wearing suits, whereas the majority wear t-shirts (we are in an experimental economics laboratory where students are employed for tests), and that there is a stain on the wall, but not *focus on* these features of the environment, nor report them. He may not even *notice* them. The scientist, finally, *reports* orally and/or in writing what (but sometimes not *all*) he has focused upon. He produces notes, tables of data, graphics, and finally *interpretations* of data. It is at this stage that the step from data to phenomena is taken, and at this stage, according to Duhem, theory comes into play.

> Open any report at all of an experiment in physics and read the conclusions; in no way are they purely and simply an exposition of certain phenomena; they are abstract propositions to which you can attach no meaning if you do not know the physical theories admitted by the author. When you read, for example, that the electromotive force of a certain gas battery increases by so many volts when the pressure is increased by so many atmospheres, what does this proposition mean? We cannot attribute any meaning to it without recourse to the most varied and advanced theories of physics (Duhem, 1906, pp. 147-8).[139]

Surely the very language used to draw the conclusions in a scientific paper does involve theoretical presuppositions: how could we speak of 'currents' and 'potentials' without some theory of electricity, of 'first order stochastic dominance' or 'risk-averse behaviour' without some theory of choice under

---

[138]Notice that this is *not* the sense in which Hanson uses the term: for him 'to see' means to receive some stimuli and a*t the same time* to interpret them.

[139]It must be noticed that this is one of the few passages where Duhem seems to give a semantic flavour to his famous thesis of underdetermination, as Quine would explicitly do in 1953.

uncertainty? Duhem, however, was not ready to go further than that. He was not willing to question the purity of sense-data, in particular:

> When a sincere witness, sound enough in mind not to confuse the play of his imagination with perceptions, and knowing the language he uses well enough to express his thought clearly, says that he has observed a fact, the fact is certain [...].
> [I]f the physicist restricts himself to narrating the facts he has seen, in the strict sense of seeing with his own eyes, his testimony should be investigated in accordance with the usual rules for determining the degree of credibility of the testimony of a man (Duhem, 1906, pp. 158-9).

The neutral character of empirical evidence must then be corrupted at some stage between the report of what has been perceived (the readings of values on a thermometer's scale, for instance) and the conclusions drawn from them (that the temperature was such-and-such). There is no doubt that Duhem's theory-ladenness thesis was the one classified under *e* above. Henri Poincaré (1905) used to speak of 'crude facts' in a sense similar to Duhem's 'practical facts', as the neutral realm of everyday experience common to all sane human beings. The post-positivists who revived his argument held more radical views. Thomas Kuhn, for instance, seems initially to be restating Duhem's point:

> Looking at a contour map, the student sees lines on paper, the cartographer a picture of a terrain. Looking at a bubble-chamber photograph, the student sees confused and broken lines, the physicist a record of familiar subnuclear events (Kuhn, 1962/1970, p. 111).

Kuhn seems to suggest that both the layman and the expert *see* lines on paper or on the photograph, but only the latter can make sense of them 'scientifically' (rather than, for instance, artistically). Shortly after, however, such a claim is rejected: "what occurs during a scientific revolution is not fully reducible to a reinterpretation of individual and stable data" (1962/1970, p. 121). Kuhn makes two points: first, two scientists guided by different theoretical presuppositions *focus on* different aspects of the same object under study. The Aristotelian facing a swinging stone "would measure (or at least discuss - the Aristotelian seldom measured) the weight of the stone, the vertical height to which it has been raised, and the time required for it to achieve rest". Galileo, on the other hand, "measured only weight, radius, angular displacement, and time per swing, which were precisely the data that could be interpreted to yield Galileo's law for the pendulum" (Kuhn, 1962/1970, pp. 123-4).

Secondly, contrary to Duhem and the neopositivists, Kuhn did not believe in the existence of 'pure' observation. He took paradigms to be so pervasive to affect the very conceptual framework according to which we classify things. Hanson (1958) had already denied that the act of receiving some stimuli could be distinguished from that of organising them in an interpreted framework. Similarly, to Kuhn "the scientist who looks at a swinging stone can have no experience that is in principle more elementary than seeing a pendulum. The alternative is not some hypothetical 'fixed' vision, but vision through an alternative paradigm, one which makes the swinging stone something else" (1962/1970, p. 128).

Kuhn's version of the theory-ladenness argument includes surely *b* and *c* above, but probably also *a*: in later writings, Kuhn noticed that although the distinction between stimuli and sensations (what stimuli we are presented with and what we 'observe') may seem at first sight plausible, it is not true that there is a 'correct' sensation for each stimulus, and therefore since "members of different communities are presented with different data by the same stimuli", "the given world, whether everyday or scientific, is not a world of stimuli" (1974, p. 309 n.). Paul Feyerabend provides the most radical statement of the thesis:

> There are not two acts - one, noticing a phenomenon; the other, expressing it with the help of the appropriate statement - but only one, viz. saying in a certain observational situation, 'the moon is following me', or 'the stone is falling straight down'. [...]
> This unity is the result of a process of learning that starts in one's childhood. From our very early days we learn to react to situations with the appropriate responses, linguistic or otherwise. The teaching procedures both *shape* the 'appearance', or 'phenomenon', and establish a firm *connection* with words, so that finally the phenomena seem to speak for themselves without outside help or extraneous knowledge. They *are* what the associated statements assert them to be (Feyerabend, 1975/1993, p. 57).

Feyerabend calls Kuhn's paradigmatic assumptions 'natural interpretations': like Kantian categories,"they are instrumental in *constituting* the field [of sensations]", rather than being "added to a previously existing" one (1975/1993, p. 60). *Unlike* Kantian categories, though, they can change and one can be taught to see differently. Feyerabend surely subscribes to *a*, the strongest form of theory-ladenness.

The move leading to such a radical position consists in *conflating* the act of perceiving a stimulus with the act of interpreting it. Hanson, Feyerabend and Kuhn usually argue by means of examples from psychology, like the famous duck-rabbit picture (figure 5).

Figure 5: The duck-rabb:

In such cases, however, there does not seem to be an objective fact of the matter behind the interpretations given by the observers. The picture represents a duck if observed from a certain point of view, and a rabbit if one performs the appropriate 'Gestalt switch'. But it is *neither* a rabbit *nor* a duck: it is just a set of lines on paper.[140] The only true, uninterpreted, pre-theoretical, neutral observational report would consist in a description (and possibly a reproduction) of the pattern of lines.

Take in contrast a drawing such as the one reproduced in Figure 6 below. It represents a micro-particle event observed by the inventor of the cloud chamber, the experimental physicist Charles Wilson. The first question concerns whether Wilson's sketch is an accurate representation of what he actually *saw* on the 29th of March 1911. Secondly, one can attempt a linguistic description of this visual image, which would go more or less in the following way: 'A circle circumscribes twelve narrow lines crossing each other in various ways, while a finger-like dark area extends from the south-western edge to approximately the centre of the circle'. Upon such a low-level observational statement, one can then build a phenomenal description like Wilson's: "On one occasion in addition to ordinary thread-like rays, one larger finger-like ray was seen, evidently a different form of secondary ray - giving rise to enormously more ionisation than even ordinary [alpha] ray".[141] The observational statement may be true or false, but there seems to be little room for disagreement. Much more controversial is the *interpretation* just given, which requires a number of theoretical assumptions above and beyond the basic observational report.

[140]For such a point, see Newton Smith (1981, p. 118).
[141]Quoted in Galison and Assmus (1989, p. 261).

Figure 6: Wilson's cloud-chamber event (from Galison and Assmus, 1989, p. 262).

Lines on paper may thus constitute a solid and neutral enough bedrock upon which to construct evidential agreement. Common reports at such a rather basic level are not sufficient to settle scientific disputes - but at the same time data such as the position of a pointer on a scale, or the track on a bubble chamber picture are rarely the issues at stake in scientific controversies. Scientists usually debate at a higher level of analysis than that, and it is not clear that the radical theory-ladenness thesis can help us understanding how scientific controversies are resolved and why they are not. For philosophical and empirical reasons, the most radical form of theory-ladenness argument is also the most difficult one to refute (how to prove that I don't 'see' what you 'see'?), but we do not have to be concerned with it here. Experimental psychologists like Slovic and Lichtenstein see people *choosing* and *pricing* just like economists do, although they may not agree on what they *prefer* (or even *that* they prefer, in economists' technical sense).[142]

In the next section, I shall try to show that scientific disagreement has normally a two-fold origin: it is either reducible to a problem of (un)reliability of the observation statements, or to a problem of fallibility of the assumptions and theories needed to infer from data to phenomena. In both cases, I shall argue, the problem can be framed in Duhem-Quine form, and the disagreement can in principle be eliminated by means of further testing.

3.9. Fallibility and reliability

The first problem with observational reports has to do with the use of theoretical terms. Popper thought that even the simplest reports cannot be theory-free due to the ubiquity of what he called 'dispositional terms'. Universal names, "words like 'glass' or 'water' are used to characterize the *law-like behaviour* of certain things" (Popper, 1934/1959, Appendix *x). But of course, one does not have to know that the term 'water' corresponds to 'a compound of two atoms of hydrogen and one of oxygen', when reporting that 'the value read on the scale of a thermometer immersed in ten litres of boiling water was 97.3 degrees Celsius'. It is worth therefore to draw a distinction between theoretical assumptions, in the sense of formal, explanatory high theory, and other non rigorous

[142] My claim, to reiterate, is that in most controversial cases a neutral level of linguistic agreement can be reached. I am not, however, addressing the problem of how such a basic observational language is formed. On this generally neglected issue, see for instance Gooding's (1990) study on the history of electromagnetism.

presuppositions.[143] Following the convention introduced in the first chapter, I shall whenever relevant write ‘Theory’ (capital ‘T’) to denote the former and ‘theory’ (small ‘t’) for the latter. High Theory is rarely - if ever - involved in reporting ‘crude facts’. There are famous examples of data analysis done by laymen: for selecting the 290,000 bubble-chamber photographs taken at CERN during the experiments on weak neutral currents in 1973, non-experts were employed after a brief training.[144] None of them knew the physics of small particles. As Hacking (1983) has convincingly argued, knowledge of high theory is seldom required in order to become a good ‘observant’, i.e. a person particularly skillful at distinguishing certain patterns in a messy complex of sense-data.

What is the real problem raised by the presence of dispositional terms in observational language, then? Popper seems to be concerned with the *fallibility* of sentences including universal names: every such statement, he says, “has the character of a hypothesis” (1934/1959, § 25). Universals like ‘glass’, or ‘horse’ refer not only to the individuals to which they have been applied in the past, but also to all future cases. They therefore cannot be ‘constituted’ by experiences, as Carnap and other neopositivists wished, because of the problem of induction affecting verificationism. Was the animal we have just seen *really* a horse? We may find out tomorrow, for example, that our theory of the classification of animals is partly mistaken (take the famous case of whales - formerly classified as fishes and then as mammals). Popper thought the only solution to be a mild form of conventionalism on observational sentences: whenever they are questioned, a test must be devised to settle the dispute. Quarrels must be followed by further tests, until an agreement is reached. The regress can be stopped only by a *decision* to consider some reports as trustworthy.

The Popperian idea of an infinite regress of justification is common to other accounts of theory-testing. Clark Glymour’s (1980) ‘bootstrapping’ account of confirmation, for example, similarly admits that the relevance of a *single* body of evidence for a *single* theory be indeterminate. A piece of evidence *e* is relevant to a certain theory *T* only given a background of assumptions *A*; other evidence *e*’ can be put forward in order to test *A*, but it will be relevant only given some other background *A*’, and so on *ad infinitum*.

[143] Cf. Hacking (1983), ch. 12 for a taxonomy of different layers of theory typically involved in scientific activity.

[144] The story of this experiment is told in Galison (1987).

It is not clear, however, whether such a regress is a vicious one. To shift the question of artefacts towards the problem of the *fallibility* of observational statements may on the contrary be a step forward. Once we move out of the swamp of incommensurable sense-data, the problem of theory-ladenness becomes in fact more tractable. As I shall argue below, problems of fallibility can be reduced to Duhem-Quine form, and the Duhem-Quine problem can in the best cases be (partially) solved by checking each component of the inference at stake in other experimental circumstances, until every source of disagreement has been eliminated.[145] Laboratory experimentation is particularly efficient in this regard, as we shall see in section 3.10. *The problem of artefacts*, in particular, *is mainly a problem of fallibility*.

A second problem concerns the *reliability* of our observations. Scientists, as a matter of fact, do disagree on observation reports. Their disagreement, however, is rarely if ever a question of radical perceptual incommensurability. Most often, scientists doubt the correctness of others' reports, and look for mistakes in their data. For instance, the problem of data in the form of 'scintillations' (tiny flashes) counts was at the core of the famous Vienna-Cambridge controversy on protons and alpha-particles in the twenties.[146] The parties disagreed upon whether 'normal' observers could count the correct number of scintillations occurring during an experiment. But, as Bogen and Woodward (1988, p. 311) point out, this was a matter of *reliability*, not of *theory-ladenness* of data. There was here a clear matter of fact at stake - i.e. the true number of scintillations - and a solution to the controversy could have been reached, if one could just devise an independent and more reliable way to ascertain the fact of the matter.

Another common source of disagreement is the quality of the data used to draw a certain inference to the existence or not of a phenomenon. During an experiment that has become a classic for philosophers of science,[147] the physicist Robert Millikan established the exact measure of the charge of an electron (a measure which is still accepted as the correct one) by ignoring a substantial portion of the data he had collected; similar data had led others (e.g., Ehrenhaft) to think that not all electrons have the same charge but instead sub-units of charge

---

[145]In some limit cases, one may be able to check each component of a system of assumptions in isolation. In that case, a sort of 'reduction' of the Duhem-Quine problem to a minimal size is operated (in a sense, it was only a pseudo-problem generated by a particular experimental procedure, which was evidently not the best possible one).

[146]See Stuwer (1985).

[147]Since Holton (1978).

exist. The issue at stake is here the *inference* from collected data to phenomena, not the theory ladenness of observation. The problem is to ascertain whether Millikan had good reasons to discard part of his data. Such reasons would, in the best case (i.e. excluding the hypothesis of fraud, a hypothesis which in Millikan's case cannot be ruled out in principle[148]) have involved some theoretical argumentation - thus shifting the burden onto the issue of the correctness of the theories involved in Millikan's data selection. To check each bit of the argument is often difficult and sometimes impossible in practice, but (and here is the important philosophical point) not in principle.

The reason why problems of fallibility and reliability are less worrying for those who want to defend the objectivity of scientific knowledge than the radical 'theory-ladenness of sense-data' argument, is that *they can be reduced to Duhem-Quine form*. It is well-known that, according to Duhem, a scientific prediction can be made only by putting to work a "whole theoretical scaffolding" (1906, p. 185). It is customary today to interpret Duhem's thesis broadly, and include among the premisses used to deduce a prediction a number of assumptions about the functioning of the instruments, the non-interferences of disturbing factors (or fulfillment of the *ceteris paribus* clause), the correct specification of the initial conditions, and so on.[149] When we seem to have produced a phenomenon contradicting our predictions, so the argument goes, we cannot *by deductive logic alone*, argue for the falsification of any one in particular of the assumptions involved (although we know that at least one must be false).

Putting it in terms of 'autopsychological' reports,[150] for instance, the prediction 'I will read on the thermometer's scale a value of 100 degrees Celsius' can be deduced from a number of assumptions including the laws of thermodynamics, the assumption that the thermometer is not damaged, that the measurement has really been made at the sea-level, that the water is pure, assumptions about the absence of major disturbances, and the proviso that I will not hallucinate when reading the thermometer. If I do not observe the predicted values, I shall not know automatically which of the assumptions to blame. But still, there will be some way to try and figure it out: by means of further testing.

---

[148] See the discussion in Franklin (1986), pp. 140-162.
[149] Cf. e.g. Quine (1953), Lakatos (1970) and Putnam (1974).
[150] Like, e.g., Watkins' (1984) 'level-0 propositions'.

Problems of fallibility and reliability of data, which can be reduced to Duhem-Quine form, are more tractable than problems of radical theory ladenness (if there exist in science any genuine examples of the latter). In principle, it should always be possible to 'extract' the relevant assumptions from the 'theory laden data' and add it as a premise to make a larger Duhem problem.[151] When the issue is the reliability of these assumptions, experiments can be devised to settle at least some disputes. The Cambridge-Vienna controversy was settled, among other things, by showing that human visual perception is unable to distinguish with the required precision different kinds of scintillations observed through the microscopes available at the time. It would have been possible *in principle* to write down the *right* report, it was just very difficult. Feyerabend and Kuhn's point applies to those cases in which there does not seem to be *one* 'right' way to see something, when there simply is no matter of fact to be ascertained. When the issue is reliability or fallibility, there is a way out of the corner: by means of successive tests we rule out some possibilities from the number of interpretations opened by the Duhem-Quine problem. In the next section I shall show how this can be done in practice by illustrating some experiments devised to show that preference reversals are a real phenomenon indeed. The problem of artefacts is a Duhem-Quine problem, and this can in principle (i.e. in particular cases) be solved by ingenious testing.

3.10. Independent tests

Neither Popper nor any other philosopher aware of Duhem's problem ever argued that a decision can be taken about which element (among the premisses involved in a predictive argument) to revise *without a reason*. The point is that such a reason cannot be an ultimate, logically compelling one. Otto Neurath (1934) noticed that *in practice* one does not have to deal with an infinite number of alternative explanations for a given scientific observation. Testing is an obvious way to try to reduce the number of plausible alternatives. The same idea can be found in Popper's writings: suppose that, as in the BDM case, someone challenges the standard inference from an observation to the falsification of a theory by putting forward a rival theory. In order to test the legitimacy of such an alternative, one has to devise an *independent test* for it.

[151] Cf. Worrall (1991) for a similar argument directed against Feyerabend's version of the theory ladenness thesis.

Popper puts forward a definition of independent test that is strictly related to his notions of ‘severe test’ and ‘non ad-hoc’ hypothesis. According him (1963), an empirical test is not *severe* if the theory at stake together with accepted background knowledge predicts just the same facts that are implied by background knowledge alone. The background includes everything that is tentatively accepted at the time the theory is tested. Evidence predicted by the theory but not by the background alone is said to be ‘independent’. An explanation is not *ad hoc* if the *explanans* (the theory) is rich in content, i.e. has testable consequences beyond the *explanandum* (i.e. the refuting instances of previous tests) at hand (Popper, 1957).

Such intuitions have inspired neo-Popperian criteria of appraisal of scientific theories: according to Lakatos (1970) a theory is non-*ad hoc*$_1$ if it predicts novel facts (Popper’s ‘richness of content’), and non-*ad hoc*$_2$ if some of these novel predictions have been corroborated by empirical facts. In order to capture the second intuition, Elie Zahar (1976) and John Worrall (1978; 1985) have proposed a rather sophisticated criterion, according to which the facts which truly corroborate a theory have to be ‘novel’, but only in a relative sense: they may be already *known* at the time of the production of the theory, but *new for that particular theory*, i.e. truly confirming phenomena must not have played a role in the construction of the theory at stake. Worrall and Zahar include among the corroborating evidence the data known at the time of the experiment, but which were not used to build the prediction at stake. Evidence cannot be used twice, in other words, once to construct and once to confirm.

We already said that Karni and Safra’s (1987) paper applies a specific generalised expected utility model - EURDP - to PR data. The way i which this is done suggests that the application is mostly illustrative. Certain passages indicate, however, that Karni and Safra believed that subjects do indeed violate independence and that this fact is the basis for an explanation of their behaviour.[152] But on what grounds?

---

[152] For example: “What Grether and Plott tried and - as our discussion indicates, *failed to do* - is to observe, by means of [the BDM method], the certainty equivalents of given lotteries” (1987, p. 676, *my emphasis*). In a footnote Karni and Safra compare their contribution to Holt’s: the latter pointed independently to violations of intransitivity, but

Perhaps Karni and Safra were influenced by the general consensus achieved by Quiggin's and Yaari's EURDP among decision theorists. Generalised theories of decision making, however, do not imply that subjects violate independence - only that they might. Such theories display in their general form several free parameters which have to be fixed in order to derive precise implications about subjects' behaviour. For some special values of these parameters, the consequences of EURDP are identical to those of expected utility theory. In other words, EURDP may well be true and yet the subjects not violate independence when choosing among the lotteries typically used in PR experiments. Some *specific models* must be employed in order to account for PR - as Karni and Safra did in order to illustrate their main result. It is clear that such models needed independent confirmations before they could be taken seriously.

The models are obtained by 'ad hoc' specification: the Karni and Safra reinterpretation of the BDM procedure holds, in fact, only for some pairs of lotteries and some values of the free parameters of the basic EURDP theory. According to the latter, the value V of a lottery (x_1, p_1;...; x_n, p_n) is given by

$$V(x_1, p_1;\ldots;x_n, p_n) = \sum_{i=1}^{n} u(x_i)\left[f\left(\sum_{j=i}^{n} p_j\right) - f\left(\sum_{j=i+1}^{n} p_j\right)\right].$$

The u is the traditional monotonic increasing real valued function defined on some interval in the real line (that is, on a range of monetary prizes). Compared to expected utility theory, EURDP has one more free parameter, namely the 'probability transformation function' f. Karni and Safra (1987) show that *if* the following specifications are chosen for f and u,

$$f(p) = \begin{cases} 1.1564p, & 0 \le p \le 0.1833 \\ 0.9p + 0.047, & 0.1833 \le p \le 0.7 \\ 0.5p + 0.327 & 0.7 \le p \le 0.98 \\ p, & 0.98 \le p \le 1, \end{cases}$$

"however, did not present an alternative theory *explicating* the 'PR' phenomenon" (p. 676, n. 4, *my emphasis*).

$$u(x) = \begin{cases} 30x + 30, & x \leq -1 \\ 10x + 10, & -1 \leq x \leq 12, \\ 6.75x + 49, & 12 \leq x, \end{cases}$$

then for lotteries such as the ones used by Grether and Plott (1979) - i.e. (-1, 1/36; 4, 35/36) and (-1.5, 25/36; 16, 11/36) - the ‘announced price reversals’ can be accounted for. (Notice: the ‘announced price reversals’ are the *data* to be explained, as opposed to the allegedly artefactual ‘PR’ *phenomenon*.) Still, these specifications are able to account only for the above lotteries: the Karni and Safra hypothesis cannot even rationalise all the data known at the time, unless one uses different parameter specifications for each experiment.[153] But even if this problem could be overcome, there would remain a general methodological concern. The illustrative *model*, with its particular parameters and initial conditions, rather than Quiggin’s *theory* is doing most of the work. EURDP cannot by itself even account for the particular asymmetries of observed reversals: only the model with its specific parameters can. The illustrative model above (theory plus specification of the free parameters plus initial conditions) was in fact devised explicitly in order to account for the evidence to be explained. The latter, then, cannot provide much support to the violation of independence hypothesis.

Following Zahar (1997), let us represent a theory with two free parameters (a_1 and a_2) as $T(a_1, a_2)$. A specific model of the theory can be devised by determining the free parameters on the basis of some empirical evidence e. Such evidence is used together with the theory T in order to deduce the values of the parameters: $[T(a_1, a_2)\ \&\ e] \Rightarrow T(a_1^*, a_2^*)$. The evidence e is in some cases, such as the present one, the very evidence the theory was intended to explain. It

[153] With hindsight, the popularity among economists of this explanation of PR appears puzzling (see also Hausman, 1992, and Hausman and Mongin, 1998): why was Karni and Safra’s work so attractive? At least three explanations can be tentatively put forward: (1) the PR phenomenon looked so damaging to orthodox theory, that economists were eager to believe more or less in *any* defensive argument whatsoever; (2) the strategy of weakening independence to account for counterexamples was then quite fashionable, therefore the Karni and Safra argument promoted theoretical unification; (3) the analysis of the BDM and RLS mechanisms had independent theoretical interest.

is not surprising, then, that e is accounted for by T.[154] Some new evidence e' not used to derive the model $T(a_1^*, a_2^*)$, but *implied* by $T(a_1^*, a_2^*)$, could confirm T more than e did, and even more so if no alternative theory is able to account for e' (which will therefore count as a quasi-crucial experiment with respect to T and its rivals).

But why do certain data provide a better test of a hypothesis than others do? According to Deborah Mayo (1996, Ch. 6), this does not have so much to do with novelty as with expectation.[155] Some data test a hypothesis severely only if it is very unlikely that those data can be produced while the hypothesis is false. The evidence e (the observed choice-price reversals) does not support Karni and Safra's model because the classical PR experiment produces results consistent with that model no matter what the real truth-value of the latter. If we repeated the PR experiment even a thousand times, it would not be surprising to find that the resulting evidence is consistent with the Karni-Safra model - because the Karni-Safra model was constructed so as to accomodate the result of a PR experiment. The role of experiments is not to produce no-matter-what data. A good experiment must produce good data to answer specific questions, better than 'casual' data would. The original PR experiment was devised to answer Lichtenstein and Slovic's question about the context-dependence of pricing and choice behaviour.[156] Other PR experiments (e.g., Grether and Plott's) were performed later in order to answer other questions about artefacts. All these tests were designed so that it would have been really unlikely that a certain result (e)

---

[154] Notice that e *weakly* confirms T, because it is logically conceivable that for some other theory $T_i \neq T$ there exists no set $\{a_1^*,...,a_n^*\}$ such that $T_i(a_1^*,...,a_n^*) \Rightarrow e$. In the limit case in which this were true for all T_i, T would be practically testable 'in isolation' and the Duhem problem would be drastically reduced. But, as in the case at hand, there usually exist a number of alternative theories able to account for e.

[155] Mayo (1996, Ch. 8) argues that her notion of severe test is able to account for all the positive intuitions of neo-Popperian theories of confirmation as well as to avoid some of their defects. The thrust of her argument against Worrall and Zahar is that there are cases in which data are used to construct a scientific hypothesis *and* at the same time provide a severe test of that hypothesis. The interested reader can look in particular at sections 8.3-8.4 of her book.

[156] See section 2 above.

was observed if the tested hypothesis were false.[157] The data produced in those experiments cannot test Karni and Safra's hypothesis severely. Some special experiment must be devised that is able to answer this new question in a convincing manner.

This reasoning can account for the behaviour of economic experimentalists, who soon began to look for genuine confirmations to the Karni and Safra hypothesis. Subsequent work by Safra, Segal, and Spivak (1990b) was devoted to deriving further testable implications from the 'violation of independence' interpretation of PR's. According to Safra, Segal, and Spivak's *Proposition 2* (derived from Karni and Safra's model), although the optimal selling price (π) of a lottery and its certainty equivalent (*CE*) may end up non-identical in a BDM elicitation, they should nevertheless lie on the same side of the lottery's expected value for the lottery (*EV*). In other words, the two following testable predictions (for risk-loving and risk-averse subjects respectively) can be derived from Karni and Safra's interpretation:[158]

(i) $CE(X) > EV(X) \Rightarrow \pi(X) \geq EV(X)$
(ii) $CE(X) < EV(X) \Rightarrow \pi(X) \leq EV(X)$

According to Segal's (1987) interpretation of the BDM device, on the other hand, (i) and (ii) do not necessarily hold. An experiment testing such predictions would certainly count as an *independent test* of Karni and Safra's hypothesis. Keller, Segal and Wang (1993) ran such an experiment, and found *Proposition 2* to be inconsistent with around 30% of the data. Such a percentage can hardly be explained as a random error, since a definite asymmetric tendency is discernible in the data: the $\pi(X) > EV(X) > CE(X)$ pattern is displayed for 22% of the subjects, whereas the $CE(X) > EV(X) > \pi(X)$ pattern is shown for a 9% only (Camerer, 1995, p. 659). This seems to rule out the explanation of PRs in terms of violations of independence put forward by Karni and Safra (1987), leaving open the issue

[157] For instance, we can formulate one of the (low-level) hypotheses tested by Grether and Plott (1979) as "*e* is not due to the absence of relevant monetary rewards". By performing experiments with relevant monetary rewards, Grether and Plott constructed a severe test of their hypothesis, by making it unlikely that *e* was produced in their experiment and at the same time the experimental hypothesis false.

[158] For the technical details of such a derivation, Cf. Safra, Segal, and Spivak (1990b, pp. 187-188). Notice that (i) and (ii) can be derived also from EUT, since according to the latter $CE(X) = \pi(X)$.

whether price reversals are effects of reduction violations (as conjectured by Segal) or symptoms of intransitive preferences.

Safra, Segal and Spivak (1990a) also proved that the kind of explanation that Karni and Safra gave of the PR phenomenon presupposes similar conditions to those known to be at the origins of other anomalies like Allais' paradox and the so-called common ratio effect. All these phenomena (including PRs) ought therefore to be explicable in Machina's Generalised Expected Utility Analysis (1982), by assuming 'fanning out' of indifference curves in a Marschak-Machina triangle as implied by Hypothesis II (i.e. that local utility functions over stochastically dominating lotteries show greater risk-aversion than utility functions over stochastically dominated ones).[159] MacDonald, Huth, and Taube (1991) devised an experiment to test the 'violation-of-independence' explanation of PRs by checking whether there is a correlation between fanning out and reversals. They ran Allais-type experiments and then PR experiments, and found that subjects who did *not* exhibit fanning out did *not* in fact incur fewer reversals than the others.

Another example of the same kind is provided by Starmer and Sugden's (1991) attempt to test Holt's explanation of PRs. Starmer and Sugden designed an experiment in which the reduction hypothesis, upon which Holt's explanation is built, is incompatible with a very frequent violation of independence first discovered by Allais (1953/1979), the common consequence effect. The Allais-type experiment involved a double choice, first between a lottery *R'* = (£10, 0.2; £7, 0.75; 0, 0.5) and a lottery *S'* = (£7, 1); then between *R''* = (0, 0.8; £10, 0.2) and *S''* = (£7, 0.25; 0, 0.75). The common consequence effect is a tendency to choose *S'* > *R'* and *R''* > *S''*. By reduction, it is easy to show that the following equivalence holds between compound lotteries: (*R'*, 0.5; *S''*, 0.5) = (*S'*, 0.5; *R''*, 0.5) = (£10, 0.1; £7, 0.5; 0, 0.4). If there is reduction, then, one should expect a random pattern of choice between (*R'*, 0.5; *S''*, 0.5) and (*S'*, 0.5; *R''*, 0.5) whereas common consequence implies (*S'*, 0.5; *R''*, 0.5) > (*R'*, 0.5; *S''*, 0.5). If there is reduction, in other words, there cannot be common consequence effects, and vice-versa. Starmer and Sugden performed the above Allais-type experiment with and without the RLS mechanism, and observed the same ratio of common consequence violations in all cases. This provided strong evidence that subjects did not obey reduction, and the Holt explanation was discredited.

[159]See chapter 2, section 4, for more details.

These experiments have reinforced scientists' belief in the reality of the preference reversals phenomenon *by elimination of alternative explanations*. Attempts to explain away the phenomenon have been rejected by testing the various alternative hypotheses independently. In all these cases, the logic of experimenting seems to follow the classic logic of theory-testing theorised by the supporters of the hypothetico-deductive model: a theory is proposed; predictions are derived from it (plus initial conditions and auxiliary hypotheses); finally, such predictions are checked against empirical data, and either the assumptions are rejected, or they are confirmed. There are however other means to increase our belief in the reality of a phenomenon or to reduce the plausibility of artificiality claims. These have to do not so much with the classic scheme of hypothesis testing, but rather take the form of 'observations without (a unique) theory'. I shall turn to these in the next section.

3.11. Phenomena without (a unique) theory

Most of those who have claimed to have responded to Hanson, Kuhn and Feyerabend's challenge have pointed to a solution of Duhem's and Popper's fallibility (or reliability) problem. Ian Hacking, in particular, devotes a good deal of his *Representing and Intervening* (1983) to defining boundaries to the importance of theory. The issue at stake in the experiments on preference reversals did not have primarily to do with any explanatory Theory (capital 'T'). It can rather be summarised in the question: how can we know that the observations of preference relations made *via* elicitation mechanisms were reliable?

Paul Feyerabend built his case for Galileo 'the propagandist' on a number of arguments. In *Against Method* (1975/1993), Feyerabend argued that a 'neutral' observer did not have a good reason to believe in Galileo's observational claims rather than in those of his opponents, because Galileo had no theory of how the telescope used to observe the moons of Jupiter worked. In the absence of such a theory, Feyerabend says, Galileo had to use rhetorical tricks in order to convince a (justifiably) sceptical scientific community. *If one does not have a theory of the instruments*, in other words, *one is not entitled to believe in what the instruments show*. It is surprising to find such an argument right at the end of a chapter entirely devoted to showing what kind of problems the pervasiveness of theory raises to 'objectivist' philosophies of science. It is an argument, however, which can often be found in the mouths of scientists. It once again dates back at least to Duhem, who suggested that theories of the instruments enable one to tell artefacts from real phenomena:

> To what strange errors we should be exposed at times, if we naively attributed to the observed objects the shape and colour revealed by the instrument, or if a discussion of optical theories did not allow us to distinguish the role of appearances from that of realities! (Duhem, 1906, p. 154).

Echoing Duhem, Edi Karni and Zvi Safra, in their paper devoted to challenging the BDM elicitation mechanism, pose the following two questions: (a) "How rich is the class of preferences that permits the elicitation of certainty equivalents of given lotteries using [the BDM] method?"; and (b) "Are there experiments that enable the elicitation of the certainty equivalents of every lottery for every reasonable preference relation?". The first question is the one that motivated their enterprise; the second one is rather more ambitious. Of course they are legitimate questions, and indeed interesting ones from a scientific (and *theoretical*, in particular) point of view, but their relevance to the issue at stake (i.e. the artefactual nature of PRs) is far from clear.

The answers provided by Karni and Safra are respectively that "(a) the elicitation of certainty equivalents of all lotteries, using the experimental methods of Becker, DeGroot, and Marschak, is possible if and only if the preference relation is representable by an expected utility functional; (b) every experiment in a larger class of experiments [which Karni and Safra call 'Q-experiments'] would fail to elicit the certainty equivalent of some lotteries for some reasonable preference relations" (Karni and Safra, 1987, p. 676). In other words, if subjects' decisions violate independence in the cases at hands, then the BDM procedure and similar mechanisms are not adequate instruments to determine certainty equivalents *in a precise way and in all cases*. From this, Karni and Safra conclude that "Grether and Plott and others [...], as our discussion indicates failed to [...] observe by means of an experimental method developed by Becker, DeGroot, and Marschak (1964), the certainty equivalents of given lotteries" (*ibid.*). Apart from the fact that - given what we have seen in the previous section and as later research on preference reversals has shown - this is quite clearly an overstatement, the point is that such an argument is not in itself sufficient to challenge the existence of PRs. To begin with, the first premiss (that agents really violate independence in these particular cases) had not been proven. Secondly, even if it had been proven, it does not follow logically that preference reversals cannot be observed by means of ‘Q-experiments’. The BDM mechanism and similar methods may not be *absolutely* or *generally* precise, but still *precise enough* to observe preference reversals. One natural way to see whether this is the case or not is to try to observe reversals with and without the BDM procedure, and check whether it makes any difference. Let us see more precisely how this can be done.

The tests examined in the previous section were devised to discriminate between rival interpretations of the PR phenomenon. They aimed at validating or refuting one of such interpretations, their perspective was one of struggle between theories. But remember the target of Holt, Karni and Safra, and Segal’s arguments: they intended to show that it was not intransitive preferences that experimentalists had observed in their experiments with the BDM procedure. In order to reject Karni and Safra’s, and Segal’s interpretation, therefore, one must not necessarily show that their *Theories* (capital ‘T’) are erroneous. It should be sufficient to show that it was really a feature of preferences that was observed in the experiments in question. Hacking, in a chapter entirely devoted to discuss the reliability of vision through microscopes (1983, ch. 11), argues that powerful support to the belief that what we see through electron microscopes is real is provided by the fact that the same structures are observed through light microscopes. The intuition behind this inference is captured by a so-called ‘no-miracles’ argument:

> Two physical processes - electron transmission and fluorescent re-emission - are used to detect the bodies. These processes have virtually nothing in common between them. They are essentially unrelated chunks of physics. It would be a preposterous coincidence if, time and again, two completely different physical processes produced identical visual configurations which were, however, artifacts of the physical processes rather than real structures in the cell (Hacking, 1983, p. 201).

According to such an argument, evidence obtained via independently working instruments provides strong support to the existence of a phenomenon. Our belief in the *reality* of a phenomenon, notice, can be totally independent of the *explanations* we give of such a phenomenon. We may not know the causes of the phenomenon, nor have an established theory of the instrument, and yet believe in the phenomenon and in what we see through an instrument.

Wesley Salmon (1984, p. 216) provides another example: Jean Perrin's determination of Avogadro's number through studying Brownian motion. In order to be sure that he had found the true value, however, Perrin checked his results by measuring it in alternative ways. In his 1913 book, *Les atomes*, Perrin reports thirteen different independent methods to ascertain Avogadro's number. The 'miraculous' convergence of all measures is taken to be an extremely strong proof that the result obtained was not an artefact of the procedures he had used.[160] As Allan Franklin (1986, pp. 131-135) points out, the *Review of Particle Properties* provides detailed information about the devices used for measurement. The latter may include automatic spark chambers, counters, electronic combinations, emulsions, hydrogen bubble chambers, missing-mass spectrometer, xenon bubble chambers, cloud chambers, propane bubble chambers, spark chambers, wire chambers, bubble chamber plus electronics, and freon bubble chambers. The reliability of a measurement is a function of the number of different techniques delivering consistent results.[161]

In the case of preference reversals, the phenomenon had been observed right from the beginning *with and without* elicitation mechanisms. Of Lichtenstein and Slovic's early tests (1971), only two involved the BDM procedure, but reversals were produced in all tests. This fact should have already been a puzzle to the Holt-Segal-Karni-and-Safra explanations. Years later, James Cox and Seth Epstein, two economists at the Arizona Experimental Lab and De Paul University, looked for a way to reproduce preference reversals with incentive mechanisms but avoiding possible problems with the BDM procedure:

> [...] it was necessary that we not use the BDM price elicitation procedure. Furthermore, we concluded that Karni and Safra's Theorem 2 makes it highly unlikely that anyone will be able to design a price elicitation mechanism for choices in a lottery space that does not require the independence axiom. Therefore, we concluded that it would be impossible for us to elicit true selling prices in an experiment that is designed in such a way that behavioral inconsistencies with the independence axiom are not

[160] Even a conventionalist like Poincaré was struck by such a result; see Nye (1972) for the full story.

[161] The normative force of no-miracles arguments is highly debated by philosophers of science: such arguments are to begin with *ampliative*, thus facing standard problems of induction. Moreover, they are special instances of so-called '*inferences to the best explanation*'; some criticisms applying to the latter can therefore be turned towards no-miracles claims. It would be too long here to discuss these philosophical problems - my purpose, it is worth stressing once again, being mainly descriptive in character: these are the arguments scientists use to settle controversial cases.

> confounded with more fundamental inconsistencies with decision theory. But preference reversals are inherently properties of inconsistent orderings. The absolute magnitude of prices is basically irrelevant; it is the fact that the less preferred lottery is given a higher price that represents an inconsistency with decision theory (Cox and Epstein, 1989, p. 412).

Cox and Epstein managed to create an incentive procedure able to elicit orderings without creating compound lotteries. First, they asked subjects to state their lowest selling price for both lotteries in each pair at the same time; the lottery with the lower price was then paid a fixed sum, whereas the other was played out for money. The prices were then compared to subjects' pairwise choices on lotteries obtained by reducing the payoffs of the original lotteries by the announced selling price (so that the probability distribution of returns was kept constant). The procedure is problematic because - as Cox and Epstein (1989, p. 422) conjecture - the subjects might interpret the pricing task as a choice task. Hausman (1992, p. 139), indeed, suggests that Cox and Epstein's might not even be classifiable as a genuine PR experiment. For our purposes, however, the general strategy is what matters: Cox and Epstein's procedure does not prevent the subjects from stating a higher selling price than their true reservation for the preferred lottery, but is supposed to ensure that the latter be assigned the highest price. Because of the structure of the experiment, it was not possible to control for wealth effects and for portfolio effects at the same time; Cox and Epstein decided to control for the latter and then cope with wealth effects by means of data analysis. Preference reversals were observed; the patterns of reversals, however, were quite different from those observed in previous classic preference reversals experiments. Many unpredicted reversals and fewer predicted asymmetries occurred, thus warranting the suspicion of - at least some element of - random choice behaviour (or rather confirming Hausman's scepticism!).

Amos Tversky, Paul Slovic and Daniel Kahneman (1990) devised an incentive mechanism that was intended to improve on Cox and Epstein's design. The basic idea, again, was that ordering, rather than the elicitation of true selling prices is what matters for the PR phenomenon. Rather than doing things concurrently, the subjects were *first* asked to price the lotteries in each pair seperately; and *then* were they faced with the choice task. Subjects were told that only one lottery among the highest priced and the chosen one would have been randomly selected and played. One can attempt to explain away the observed reversals by means of a generalised expected utility model assuming a mixed strategy on the subjects' part, i.e. by supposing that agents prefer a 50%

chance to play either the highest or the lowest valued lottery to the option of playing one of them for sure. Such an explanation, however, cannot account for systematic patterns such as those observed in classic PR experiments and replicated by Tversky, Slovic and Kahneman.

Finally, some experimental strategies involve a more substantial use of theory in supporting the reality of a phenomenon - but do so in order to justify the adoption of the BDM machinery rather than to explain the phenomenon. MacDonald, Huth and Taube's (1992) results suggested that a so-called 'isolation effect' may have been present in their experiments. Subjects, in other words, seemed to choose as though they evaluated each lottery in isolation, without multiplying its chances for the probability of its being selected by the BDM procedure. Such a procedure would be consistent with theories of decision, like Tversky and Kahneman's 'Prospect Theory' (1979), which assume a desire on subjects' part to minimise the computational costs of decision making. MacDonald, Huth and Taube devised an experiment where subjects were offered the chance to revise their choice after an RLS was performed at the end of the experiment. A subject not obeying independence should change his choice, but very few did change despite the fact that strong independence violations were observed in previous experiments. Thus, some isolation effect may have been there. Isolation effects have the interesting property of counterbalancing the effect of independence violations: the BDM procedure, in fact, fails to elicit true certainty equivalents only if independence is violated *and* reduction is obeyed. If subjects violate *both* independence *and* reduction by isolation, then the BDM machine may work well (as suggested by Camerer, 1989).

3.12. The reality of reversals

Colin Camerer, in a recent survey of the PR experiments, concludes that the PR phenomenon can hardly be considered an artefact of the instruments of observation. This is today the standard view. Of the arguments he cites in its support, two are from refutation of alternative explanations: (a) from failed predictions derived from generalised expected utility models; and (b) from the evidence of 'isolation effects'. The third argument (c) starts from the recognition that the same phenomenon seems to be observable *via* different mechanisms relying on different principles (Camerer, 1995, p. 659). They may not know exactly how the BDM mechanism works (although we surely understand it better now than ten years ago), but experimental economists are confident today that it may be used to observe preference reversals. Experimentalists rely on several resources to establish the reality of a phenomenon. Allan Franklin (1990) lists nine possible strategies adopted by scientists in order to provide "reasonable belief in the validity of an experimental result" (1990, p. 104):

1. Experimental checks and calibration, in which the apparatus reproduces known phenomena.
2. Reproducing artifacts that are known in advance to be present.
3. Intervention, in which the experimenter manipulates the object under observation.
4. *Independent confirmation using different experiments.*
5. *Elimination of plausible sources of error and alternative explanations of the result.*
6. Using the results themselves to argue for their validity.
7. Using an independently well-corroborated theory of the phenomena to explain the results.
8. Using an apparatus based on a well-corroborated theory.
9. Using statistical arguments.

The strategies we have been concerned with in this reconstruction of the debate on PRs belong to the italicised categories four and five above. Thanks to such strategies, the debate on the existence of PRs was in effect brought to an end, and disagreement eventually more or less eliminated. This, it must be stressed, is quite a remarkable event in economics, where controversies tend to last for decades without parties ever converging on a common position.[162] Such a result was made possible by the use of controlled experimentation. Other strategies in Franklin's list were clearly not available to economic experimentalists in the case at hand: the phenomenon under study, for instance, was inconsistent with accepted economic theory, which therefore could not be used to increase economists' confidence in the reality of PRs. The theory from which Slovic and Lichtenstein predicted the occurrence of reversals, being incompatible with the very idea of a preference structure, could not be accepted by economists. The theory of the apparatus was undermined by Allais-type violations since the early fifties. No artefacts were known to be present in advance.

3.13. Replication and Reproduction

[162] Hausman and Mongin (1998) try to address the question *why* this is so.

The reality of PRs has been (temporarily, perhaps) accepted. One does not have to have a theory of light refraction to believe that a telescope accurately reports what is there. Similarly, the *causes* of preference reversals have not been established yet, and recent investigations invite the thought that reversals may be the effect of the interaction of a number of factors at the same time - but PRs are very likely to be real.[163] Replication has been a fundamental tool in the course of the controversy. This claim may appear inconsistent with economic experimentalists' complaints about their own habits: experiments, they say, are rarely if ever replicated. The complaint usually takes the form of a critique of economic journals' standards, which tend to provide disincentives for checks and replications of others' experiments.[164] If we take a look at the natural sciences, however, we'll find that the situation is not that different: replications are rare, official scientific rhetoric notwithstanding. Harry Collins (1975) has convincingly shown that scientists rarely if ever check other experimenters' claims by repeating exactly their procedures. Given that I have begun by praising replicability as a fundamental advantage of laboratory experimentation above field data analysis, there is here a puzzle to be solved.

Mulkay and Gilbert (1986) have argued that scientists use the term 'replication' in two different senses: (1) 'mere' or 'exact' replication of someone else's experiment, to check his procedures; and (2) replication 'through experimental variation'. More importance is usually attributed to the latter activity, because "the use of various independent methods strengthens confidence in the conclusion by showing that it does not depend solely on one specific kind of experimental manipulation" (Mulkay and Gilbert, 1986, p. 27). It is useful to introduce some terminology to distinguish the two activities. Following Cartwright (1991), I shall speak of 'replication' and 'reproduction', respectively, to denote the activities (1) and (2) above.[165] PRs have in this sense now gone through both stages: the stage of replication and that of reproduction.

'*Replicability*' is a property of experiments. Using once again Bogen and Woodward's (1988) terminology, we may see an experiment as a data-producing device. An experiment becomes replicable when a group of scientists achieves a stabilization and a formalisation of the techniques to produce certain

[163]On the possible causes of PRs, see Camerer's survey (1995). Cf. also Tversky, Slovic, and Kahneman (1990) and Slovic (1995) for a statement of psychologists' point of view on the matter.

[164]Cf. Smith (1994) for a similar point.

[165]The definitions are of course partly arbitrary. Radder (1996) for instance uses exactly the opposite terminology.

patterns of data in a reliable way, when the interaction between experimenters and machines has become such that data with certain characteristics can be produced at will. In a replication the same 'data generating process' as the one used in the original experiment is at work (up to spatio-temporal differences: the experiment is possibly replicated by another team of scientists, in another laboratory, and surely at a different time). '*Reproducibility*', on the other hand, is a property of phenomena. A phenomenon has been reproduced when experimenters have become able to observe it or produce it in a number of independent ways, with the use of different instruments. Different data-sets (that is, data produced by different 'data-generating processes' or 'causal set-ups') sometimes together with their different theoretical interpretations converge to prove the existence of a certain phenomenon.

Experimenters in all sciences usually do not replicate others' results, unless there is some serious worry of incompetence or the data seem 'too strange to be true'. From this respect, then, economists do not differ at all from other scientists. Their worries are just an effect of their being prone to the rhetoric of the natural sciences, and of the ambiguity of the term 'replication'. One important philosophical issue is whether strategies such as reproduction through variation provide *rational* support to scientists' belief in a result.[166] Collins (1984) has argued that the practice of reproduction, far from being enforced by its intrinsic rationality, is an effect of sociological incentives (such as 'publish or perish').[167] But this is not necessarily a criticism nor a challenge to the rationality of science. In certain cases publishing standards can be seen as institutions evolved to stimulate an efficient and rational pursuit of scientific research. Experimental variation is an efficient mean to serve an epistemic goal: it makes the construction of alternative causal stories extremely difficult.

3.14. Conclusion

Many students of the natural sciences have noticed that experimenters are often concerned with establishing the existence and the conditions of reproducibility of some phenomenon, rather than with its theoretical explanation. In this chapter, I have tried to show that the same is true in experimental economics as well, and to

[166] See footnote 26.

[167] Collins' paper is a reply to Franklin and Howson's (1984) attempt to explicate the procedure of replication/reproduction by means of a Bayesian analysis.

illustrate some of the techniques used to warrant an experimental result. I suggest that the possibility of eliminating alternative explanations and of using different tools to elicit the same phenomenon distinguish laboratory practice from other empirical branches of economics. When data are collected in the field, the inferences drawn from them are much more questionable. Even some basic macroeconomic measurement which are normally taken for granted for policy purposes - such as those of inflation or stage of the business cycle - seem to be built on very shaky grounds.[168] Yet, the experimental economist must face up to a trade-off. The laboratory data my be more reliable, but there is a clear and major impediment to the direct *use* of laboratory results for intervention: before policy conclusions can be drawn, it is necessary to demonstrate that the knowledge gathered in the laboratory can be generalised to the 'real world' economies. How - if at all - can this be done? In the next chapter I shall state the problem in a more precise fashion and then discuss in detail how it can be tackled.

[168] For a historical study of the 'construction' of business cycles, see Epstein (1999); I have learned about the problems of measuring inflation from Reiss (unpublished).

Chapter 4

The Problem of ‘Parallelism’

Some conceptual analysis and an application

‘And what about the carpenter? Didn’t you agree that what he produces is not the form of bed which according to us is what a bed really is, but a particular bed?’
‘I did.’
‘If, then, what he makes is not “what a bed really is”, his product is not “what is”, but something which *resembles* “what is” without *being* it. And anyone who says that the product of the carpenter or any other craftsman are ultimately real can hardly be telling the truth, can he?’
‘No one familiar with the sort of arguments we’re using could suppose so.’
(Plato, *The Republic*, Part 10, 597a)

4.1. Introduction

As noticed by Peter Galison, when scientists ‘do not see how to make a phenomenon go away’ (1987, p. 235) that phenomenon is accepted as ‘real’. Of course, and crucially, there are various epistemic constraints on the strategies considered legitimate in explaining phenomena away. In experimental economics such strategies fall into either one or another of two main categories. ‘Explanations’ of the first kind have been at centre stage in the last chapter: there, I have mainly focused on the resources experimentalists use in order to reduce disagreement about the meaning of data, or in other words to tell phenomena from artefacts of the experiment. Artefacts of that kind, I have tried to argue, are a product of scientists’ incapacity to interpret data correctly, rather than a real characteristic of the object studied. But there are also artefacts of a different kind. According to the *Oxford English Dictionary*, an ‘artefact’ is

> Something observed in a scientific investigation, experiment, etc. that is not naturally present but originates in the preparative or investigative procedure or extraneously.

This definition highlights a crucial ambiguity in the notion of artefact. The 'contrast class' of 'artefacts' (in the sense seen in the last chapter) is 'phenomena'. A phenomenon is surely *real* - otherwise, it would not be a phenomenon at all, just a wrong inference from data (an 'artefact'). Phenomena may not occur spontaneously, i.e. when the situation is not constructed so as to have the 'right' conditions, but this does not affect their reality. In contrast, we have seen in the last chapter that scientists often speak of artefacts as 'illusions': for instance, that whereas announced price reversals are real, preference reversals are not.

But according to the *Oxford English Dictionary*, the first definition of 'artificial' is "made by or resulting from art or artifice; constructed, contrived; not natural (though real)".[169] Of course such a definition is not entirely satisfactory either: men are, after all, part of nature, and so are the objects they produce. The dichotomy 'spontaneous vs. forced' is just as questionable as the 'natural vs. artificial' one: no phenomena occur spontaneously in the economy, because they are the effect of the intentional actions of a number of human beings.[170] What therefore does 'artificial' mean in this case? The experimental economist Vernon Smith suggests the following account:

> Once replicable results have been documented in laboratory experiments, one's scientific curiosity naturally asks if these results also apply to other environments, particularly those of the field. Since economic theory has been inspired by field environments, we would like to know, if we were lucky enough to have a theory fail to be falsified in the laboratory, whether our good luck will also extend to the field. Even if our theories have been falsified, or if we have no theory of certain well-documented behavioural results in the laboratory, we would like to know if such results are transferable to field environments (Smith, 1982, p. 267).

---

[169]Cf. Hacking (1988, p. 285) for a discussion of the notion of 'artefact' which follows a similar line of reasoning.

[170]The final result may of course not correspond to any of the individual agents' intentions (for instance, the price in a perfectly competitive market, according to neoclassical theory, cannot be determined by any of the agents in the market), but is still not 'spontaneous' in the sense of occurring independently of human action.

According to Smith,[171] economic theory is mainly concerned with a certain realm of social phenomena, namely market phenomena. The latter mostly occur in environments that are fairly different from the experimental situations where phenomena like preference reversals have been first produced and then studied. Such experimental environments are 'artificial' in the sense of being created especially for a certain scientific purpose. They do not clearly belong to the intended (or at least the traditional) domain of economic theory. It is therefore important to check whether a claim about some experimental phenomenon can be generalised to situations lying outside the (laboratory) domain where the phenomenon has first been observed. If it does not, the phenomenon is 'artificial': real (as all phenomena are) but produced in the laboratory. Another way (perhaps more familiar to philosophers of science) to put it is this: scientific generalisations can be put in conditional (or causal) form - *if* certain conditions are instantiated, *then* some phenomenon will occur.[172] One then naturally asks whether circumstances 'sufficiently similar' (sufficiently for the instantiation of the phenomenon) to those obtaining in the laboratory are likely to occur 'naturally' (i.e. non deliberately).

In order to clarify the terminology, we may speak of artificial phenomena as either 'artefacts$_1$' or 'artefacts$_2$'. Artefacts$_1$ are misinterpreted data, mere 'illusions of phenomena'. Artefacts$_2$ are phenomena, but may be artificial nonetheless in the above sense.

	real	unreal
instantiated in the intended domain of economics	genuine economic phenomenon	artefact$_1$

[171] And to a very distinguished tradition in economics: cf. e.g. Mill (1836), Pareto (1907) and, for a contemporary formulation of the same view, Hausman (1992).

[172] This applies also to phenomenal claims not grounded (yet) on any Theory (capital 'T'). For instance, in the case of preference reversals, one can define the phenomenon and add a rough description of the experimental conditions under which it can be produced. In other words, 'if...then...' statements should not be necessarily identified with formal, rigorous, explanatory laws.

not instantiated in the intended domain of economics	artefact$_2$	artefact$_1$

For instance, preference reversals, according to Holt, Segal, Karni and Safra, do not exist at all; they are 'illusions' produced by a mistaken theory of the instruments of observation (lower right-hand box). The experiments reviewed in the last chapter have shown that this is very probably not the case (left-hand boxes); but they have not shown that preference reversals are phenomena falling in the 'intended domain' of economics (upper right-hand box). The preference reversal phenomenon may be 'non-genuine' just like an artificial heart is not a real heart (lower right-hand box). It is not an artefact of the instruments of observation, but it may be an artefact of the experiment in general: the phenomenon does occur in the laboratory, but does it occur in 'real' markets? In this chapter, I shall try to show how such a question may be answered.

Economists are particularly aware of the importance of overcoming this 'second-level' artificiality test before acknowledging experimental results. Indeed, they have created a special terminology for dealing with this issue, and often speak of the problem of '*parallelism*'. Parallelism is, at least from a historical point of view, the major philosophical problem of experimental economics: the one accompanying the field right from its beginning, and to this day the most debated one (it is debated in other sciences, of course, but perhaps not with same acrimony as in economics[173]). In this chapter I shall mainly be concerned with trying to define the problem of parallelism in a clear, coherent and useful way. *What is* 'parallelism', and *why* does it raise a 'problem'? A historical answer to these questions is as follows.

Long before the term 'parallelism' was invented (by Vernon Smith, 1976; 1982), experimenters were confronted with criticisms of the following sort:

> These papers are undoubtedly interesting and remarkably ingenious ones. Probably they are useful additions to knowledge in such fields as psychology [...], but their relevance within economics is not convincingly established. One suspects that the economists of the next generation will

[173]Other scientists, e.g. experimental psychologists, speak of 'internal' vs. 'external validity' of the experiment. More on this terminological distinction later (see section 9 below). Physiology and experimental medicine is another field where problems of parallelism are particularly relevant (cf. e.g. La Follette and Shanks, 1996).

> view such discussions in the same light in which they view the more exotic disputations preceding Niceae. The last paper in particular [Edwards, 1953], one believes, though admirable in many ways, is essentially misdirected. It makes use of the experimental method, a technique that has, perhaps, been neglected by the economist. But *the facts that it describes and discovers, the experiments that it summarizes, are surely of negligible significance within economics*. Of what possible concern to the economist is the conduct of Harvard students in gambling small sums on a crooked pin-ball machine? True, their conduct requires a cardinal measure for its summary. But this does not mean that any important aspect of consumer or entrepreneurial behaviour is thereby explained (Weldon, 1953, p. 350, *emphasis added*).

This is the transcript of a comment read at a meeting of the Econometric Society in the early fifties. The target is a paper reporting the results of some experiments on expected utility theory by the psychologist Warren Edwards (1953). It is most interesting that this is *all* Weldon had to say about Edward's experiments: that laboratory experimentation of the kind done in those years by several economists and psychologists was useless for the sake of economic knowledge. Instead of engaging in a full historical reconstruction and discussion of the debate on parallelism, however, I shall first and foremost try to analyse the two terms of interest: What can 'parallelism' mean? And what kind of 'problem(s)' can it raise? Mine will be a characteristic exercise in 'explicating' a concept which happens to be used in an inconsistent, unclear, and sometimes confused fashion by scientists.[174]

The plan of the chapter is the following: in the first part I shall try to tackle the above questions from a purely conceptual point of view, and formulate a number of different theses about parallelism. Whenever possible, I shall relate my abstract discussion to what the economists actually engaged in the debate on parallelism have said about it, and try to fit their claims into the categories I have proposed. Section 4.2 contains a preliminary discussion of this issue; section 4.3 reviews some metaphysical versions, and section 4.4 some methodological versions of parallelism. At the end of section 4.4 I shall identify - in the requirements of theory-completion and domain unrestrictedness of theories - a crucial principle invoked in order to justify most versions of parallelism (both metaphysical *and*

[174] A classical discussion of the methodology of 'explicating' can be found in Carnap (1950, pp. 5-6).

methodological). The last part of the chapter is devoted to criticism and application of the categories thus identified. In section 4.5 I shall argue that these requirements are too strong, and thus conclude (in section 4.6) that only an empirical version of parallelism is acceptable. The case of preference reversals will be re-examined in section 4.7 to show how in practice parallelism claims can be tested empirically. Section 4.8 contains a discussion of the preference reversals case in the light of the thesis of empirical parallelism, and section 4.9 states sharply the main conclusions of the whole analysis.

4.2. The relation of parallelism: preliminary discussion

The problem I am concerned with in this chapter can be summarised as follows: What is the bearing of *experimental results* (concerning, e.g., the truth of some economic generalisation, the existence of some phenomenon, or the measurement of some parameter) on *claims* concerning some real-world market? The issue, then, is the value of studying a set of objects (experimental markets or individual laboratory behaviour) to collect knowledge about another set of objects (real markets or individual behaviour outside the laboratory). The very use of the word 'parallelism', however, suggests that economists have often framed the issue in terms of a *relation* of some sort (a relation of 'parallelism') between the two sets of entities themselves.[175] This approach is not entirely satisfactory, and in the following sections (especially from section 4 onwards) I shall shift progressively towards formulations that reflect the former way of putting it (i.e. in terms of the relevance of experimental findings to economic generalisations).[176] For the time being, let me take as a starting point for my discussion an admittedly vague, and philosophically defective, definition of parallelism:

> [P1] Phenomena observed in the laboratory are similar (in some relevant respect to be identified more precisely) to phenomena in the real world.

The word 'parallelism' itself evokes some kind of relationship between two groups of objects. The above formulation thus provides a preliminary answer to three questions that arise naturally: What are the objects claimed to be 'parallel'? And given that the word 'parallelism' is clearly used metaphorically, What is the actual relation holding between them? And finally, Why is such a relation relevant, interesting, or problematic?

---

[175]But not always: see Smith's quote in footnote 171 below.

[176]I owe this point to Dan Hausman, who made the distinction between the two perspectives clear to me.

The first question (*What are the relata?)* is surely the easiest one to answer: the terms of the parallelism relationship are, on the one hand, laboratory economies, and, on the other, 'real-world' economies.[177] *What kind of relation is a parallelism relation*, then? Surely, a laboratory and a 'real-world' system cannot be identical: they differ in their spatio-temporal coordinates, at the very least. Parallelism is rather a relation of *similarity* between real and laboratory economies: a laboratory phenomenon may or may not have a 'parallel' one (its real-world counterpart) in a real economy, and conversely. Relations of similarity can be of many sorts, however. There can be 'commonsense' similarity (in both cases real people are engaged in *prima facie* similar activities like bargaining, bidding, or betting); similarity from the point of view of theory (the two economies share all or at least the most important observable characteristics that economic theory deems essential for a social system to be considered an economy); and structural, 'deep' similarity (the same causal principles are at work in both kinds of system). Finally, *Why is such a relation relevant, interesting, or problematic?* In order to understand *why*, we need first to examine various possible interpretations of the parallelism relation. The answer to the last question, in other words, varies with the answer to the second one. In the next section I shall start investigating these issues, by identifying and then discussing various versions of parallelism.

4.3. Metaphysical versions of parallelism

I shall begin by presenting some positive statements about parallelism, vaguely defined for the time being as a similarity relation of some sort (as in the previous section). Let me first introduce a distinction between 'nomic' and 'exhibited' parallelism. Parallelism is said to be *exhibited* when the behaviours of a laboratory and of a real-world system are both determined by the same causal mechanism or structure. There may be a mere token difference (such as the difference between *my* bike and *yours*), or a difference in scale (such as the

[177]The word 'real' has been put between scare-quotes because, as I have already argued in the introduction to this chapter, such a terminology is not only imprecise, but also misleading. It suggests that the ('artificial') systems studied by economists in the laboratory are not as 'genuine', as 'good' as those (the 'real' ones) outside the laboratory. But of course laboratory economies are just as real as any other economy. 'Natural' would not do the job either: 'real' economies are a human product just as much as experimental, 'artificial', ones. It should perhaps be better to speak of '*historically evolved*' economic systems. This is the true contrast-class: those systems that were already 'there' before experimenters built their laboratory systems. I shall nonetheless go on speaking of 'real-world' versus laboratory economies for reasons of simplicity.

difference between a professional's and a kid's bicycle), but the structure must be the same. *Nomic* parallelism asserts that the components of two systems obey the same causal laws or principles, but does not require structural similarity. Two wire-spoked wheels, a tubular metal frame, a saddle-like seat, handlebars, etc. do not necessarily make a bicycle - it depends on the way they are put together. For an economic equivalent, take the case of monetary payoffs: even granting that laboratory subjects were rational maximisers, for example, it does not follow that they will behave like real *economic* (i.e., typically, self-regarding) agents; one has to ensure that monetary payoffs are sufficiently high to render all other motives (other-regarding behaviour, for instance) irrelevant. Exhibited parallelism implies nomic parallelism, but not the other way around. Nomic parallelism implies only some counterfactual claims about exhibited parallelism: 'If it were the case that a laboratory system was arranged in such and such a way, then it would behave just like that real-world economy' - but this arrangement may never hold.[178]

In order to justify nomic parallelism, one can appeal to a doctrine about the 'uniformity of nature'. In economics, such a doctrine may take the form of a claim that human beings act according to the same principles and motivations inside as well as outside the laboratory. This does not mean, of course, that every observed laboratory phenomenon has its real-world counterpart: the conditions inside and outside the laboratory may differ enough to lead (sometimes, often, or even always) to different results. The idea is rather that when this is the case the difference can be explicated in terms of laws and principles that are valid in both domains. From the doctrine of the uniformity of nature there follows only that

> [P2.1] Parallelism is exhibited whenever the same causally relevant conditions and the same causal laws hold.

This principle recalls the old metaphysical *dictum* 'Same cause - same effect'. Should (whatever the true law says are) the relevant conditions repeat themselves exactly twice, the same events (or probability distribution over possible events)

---

[178]The distinction between nomic and exhibited parallelism is implicit (but not made clear) in Smith's account. As for the nomic version, see Smith's original formulation of the 'precept of parallelism' (a 'precept' being one of "a proposed set of sufficient conditions for a valid controlled microeconomic experiment" (Smith, 1982, p. 261)):
"*Parallelism:* propositions about the behaviour of individuals and the performance of institutions that have been tested in laboratory microeconomies apply also to nonlaboratory microeconomies where similar *ceteris paribus* conditions hold" (1982, p. 267).
Smith adds that even taking the above thesis for granted, "Which kinds of behaviour *exhibit* parallelism and which not can only be determined empirically by comparison studies" (1982, p. 267, *my italics*).

will follow.[179] The intuitive idea is that the behaviour of a certain economic system is determined by a subset of its own properties and of the properties of the surrounding environment only. Other characteristics are not causally efficient in determining what will happen. If we call the causally efficient properties 'nomic', we obtain a first sense in which the parallelism relation can be seen as a relation of 'nomic similarity' between systems. The problem of determining which properties are nomic and which not is of course open (we do not know the true laws of most, perhaps all, systems), but it is an epistemic problem we can shelve for the time being. We shall see later that experimenters have an answer to this question, and that economic theory is supposed to play a crucial role in identifying (albeit imperfectly) the nomic properties of economic systems.

Who subscribes to a metaphysical view of the parallelism relation? Vernon Smith refers indeed to the idea that "*as far as we can tell*, the same physical laws prevail everywhere" (1982, p. 267, *my emphasis*). It should be stressed, once more, that from the metaphysics of the uniformity of nature it does *not* follow that any parallelism will be actually exhibited. Laboratory and real-world systems may have the same dispositional properties without ever exhibiting similar causal structures. Suppose there is an inherent characteristic of economic laboratories invariably influencing the economic systems under study (a practically ineliminable experimenter's effect, to take a fictional example): such systems may very well obey the same laws as a real-world economy (in that *if* the experimenter's effect was eliminated, the same outcome would be exhibited), and yet never exhibit parallelism. The two economies may be, in other words, different machines (like a bicycle, a tandem and a tricycle) working according to the same general principles (those of mechanics). But the proviso ('as far as we can tell...') is even more important: the uniformity of nature thesis is supposed to be accepted as a conjecture, only until proof to the contrary has been put forward. This leads us to the fundamental problem with metaphysical doctrines. If we take metaphysics to be concerned with factual but not directly testable matters (as it is now customary to do, at least in the analytical tradition), for any metaphysical doctrine it should be possible to imagine another one incompatible with the former with the same (zero) direct empirical content. How can we choose between them?

One answer is that metaphysics has to be ultimately assessed 'empirically'. Although metaphysical doctrines cannot be tested *directly*, one can collect

[179]The amendment is needed to make indeterminism possible: 'Same conditions - same probability distribution'.

indirect evidence about them. The idea that the state of the universe is strictly determined at any instant by some laws and by its configuration at the immediately preceding state (i.e. the doctrine of determinism), for example, can be indirectly assessed in the light of past scientific achievements, by checking to what extent scientists' attempts at reducing every indeterministic explanation to a deterministic one have actually paid off. This is arguably the sense of that 'as far as we can tell' in Smith's quote. But then, parallelism becomes (albeit indirectly) a factual statement to be accepted or rejected on the basis of empirical evidence.

The second strategy consists in arguing that to endorse a certain metaphysics may be useful independently of its truthfulness. But one can think of examples of fruitful as well as of instances of misleading metaphysical doctrines. For instance, the Cartesian metaphysics of action at contact, however plausible, was an impediment to the acceptance of Newtonian mechanics, which later proved to be a much more progressive research programme. How can you tell a fruitful from a regressive metaphysics? The answer depends of course on the goals one is aiming at. We are entering here the realm of methodology, and leaving that of metaphysics. A metaphysical version of parallelism (as a positive statement) seems to lead eventually to either *empirical* or *methodological* versions, or both. I shall postpone the discussion of an empirical assessment of parallelism until the end of the chapter; in the next section I shall discuss normative, methodological versions of parallelism.

4.4. Methodological versions of parallelism

Let me return to the basic idea of nomic parallelism: an experimental and a non-experimental system are nomically parallel if and only if they feature the same dispositional properties, or obey the same counterfactual laws - '*if* it were the case that economic system ξ was arranged in such and such a way, then it would display the property *P*', where ξ is supposed to vary over the set of all laboratory and non-laboratory systems. Of course we are faced with the problem of discovering and establishing valid claims of this sort.

A set of methodological rules similar to those proposed in other sciences has been put forward by experimentalists for this purpose. As I shall try to show, such rules often *presuppose* some form of nomic parallelism without directly arguing for it, and therefore I shall discuss them under a different heading, as 'methodological' versions of parallelism. These forms of parallelism are 'methodological' in character because they usually appear in methodological *arguments*. The idea is that 'real-world' and experimental systems should, for reasons having to do with the advancement of science, be treated on a par. Take for instance the rule stating that

> [P3.1] Every counterexample obtained in testing a theory in the laboratory justifies the rejection of that theory.

In this case, experimental falsifications are said to be 'methodologically' justified. An experimental counterexample will tell us that the proposition under test is not generally valid, and this constitutes a good enough reason to reject it. Such a falsificationist version of parallelism is indeed often justified by appealing to a methodological requirement of generality and universality imposed on scientific theories. From the claim that good scientific theories should be general and universal in scope and applicability, it follows that they should be able to explain behaviour outside as well as inside the laboratory. The experimental economist Charles Plott says explicitly that "[economic] models are general models involving basic principles intended to have applicability independent of time and location" (Plott, 1991, p. 905).[180] In other words:

> [P3.1.1] Every negative result obtained in testing a theory in the laboratory informs us that the theories under test are not generally valid.

The above reasoning vindicates the relevance of a negative experimental result independently of its actual instantiation outside the laboratory. Such a simple falsificationist argument should be distinguished from another one adding to the evidence of laboratory falsification the prediction that the anomaly *will* (probably) occur in the real world (and thus parallelism will be actually *exhibited*, to adopt the terminology of the previous section). It is an inductive argument from failure in simple cases to failure in complex ones: in the simple case (the laboratory) the outcome phenomenon results from the interaction of a number of factors, let us call them $X_1,...,X_n$. In the complex case (the 'real world') the outcome results from the interaction of $m > n$ factors $X_1,...,X_n,...,X_m$. It may be the case that the claim

[180] Plott adds: "A staggeringly large number of theories exist. One purpose of the laboratory is to reduce the number by determining which do not work in the simple cases. The purpose is also to improve the models by exploring how a model might be changed to make it work better in the simple cases. General models, such as those applied to the very complicated economies found in the wild, must apply to simple special cases. Models that do not apply to the simple special cases are not general and thus cannot be viewed as such" (1991, p. 905).

(1) 'For all systems in L, if it were the case that $X_1*,...,X_n*$ then it would be the case that $Y*$'

is false (L is the set of laboratory economies in which only $X_1,...,X_n$ are active, X_i* is a particular value of a variable X_i), whereas the claim

(2) 'For all systems in W, if it were the case that $X_1*,...,X_n*$ then it would be the case that $Y*$'

turns out to be true (W is the set of 'real world' economies in which m factors are at work), because the 'extra' factors $X_{n+1},..., X_m$ have the effect (jointly with $X_1,...,X_n$) of determining the value of Y. According to the argument I am referring to, such a circumstance is extremely unlikely: given that (*ex hypothesis*) the theory under test models only n factors, unless it captures correctly the dependence of Y on $X_1,...,X_n$ in L, it will hardly be able to capture its dependence on $X_1,...,X_n,...,X_m$ - or, indeed, on $X_1,...,X_n$ in W.[181]

> [P3.1.2] Every negative result obtained in testing a theory in the laboratory identifies an anomalous phenomenon that is likely to arise in non-laboratory environments too.

The reasoning behind such a claim (let us call it the '*argument from the method of analysis-synthesis*': science should proceed by first formulating the

---

[181] Cf. Graham Loomes' statement: "if one or more of the fundamental axioms of expected utility theory fail under such [i.e. laboratory] apparently favourable conditions, there are surely grounds for questioning the power of the model as a *general* theory of individual decision-making under risk and uncertainty. If the basic axioms are substantially and *systematically* violated in these simple cases, how confident can we be about their validity in more complex cases?" (Loomes, 1989, p. 173).
Luis Wilde uses the same argument: "if an experiment includes all parameters relevant to a particular theory, and if the theory fails to predict well in the simplified setting of the laboratory, then it cannot be expected to predict well in more complex environments. [...] The experiment does not need to be 'realistic' and no presumptions need be made about its connection to more complex ('real-world') environments [in order to use laboratory experiments to reject some theories as nonsense]" (Wilde, 1980, p. 143).
Finally, John Hey's conception of laboratory experimentation as a way of performing a preliminary selection of theories seems to rely on a similar line of reasoning. According to Hey, experimental economics allows one to treat separately two crucial but distinct issues, namely
"1. that the theory is correct given the appropriate specification (that is, under the given conditions); 2. that the theory survives the transition from the world of the theory to the real world" (Hey, 1991, p. 10). A theory failing at stage 1, according to Hey, should not be allowed to enter stage 2 of testing.

laws of a few causal factors acting in isolation, and then by finding out about how they interact[182]) could be used to answer Charles Plott's '*problem* of parallelism', i.e.:

> What use are experimental results to someone who is interested in something vastly larger and more complicated, perhaps fundamentally different than anything that can be studied in a laboratory setting? (Plott, 1987, p. 193).[183]

The argument from unrestrictedness and the argument from the method of analysis-synthesis, eventually, point to the same conclusion: if a theory fails in the laboratory, too bad for the theory; a theory falsified in the laboratory should be abandoned - or at least modified, i.e. abandoned in its present form. Laboratory experiments are surely relevant, because theories are supposed to be general and must apply to the laboratory as well as to the field.

A statement symmetrical to [P3.1] asserts that

> [P3.2] Every *positive* result obtained in the laboratory supports the theory under test.

An attempted rationale for such a claim can, once again, be found in Vernon Smith's writings. According to Smith (1982), if a theory modelling the dependence of an effect on *n* variables has survived the test in a laboratory environment where those *n* factors (and no other) are at work, it should be taken as valid wherever those (and no other) factors are at work (by metaphysical stipulation - see nomic parallelism above). If other factors not modelled in the theory (and therefore not controlled by the experimenter) are at work in the real-world systems the theory is supposed to explain, then the problem lies with the theory, not with the experiment.[184]

---

[182] Cf. Mill (1836) and Pareto (1907) for a presentation and defence of such a heuristic procedure in economics.

[183] The way the question is posed (by asking 'what use...'), as well as Plott's subsequent discussion, suggest that his problem is not just epistemic, but more in general pragmatic. I shall here focus on the problem of epistemic validity only.

[184] "If [an experiment's] purpose is to test a theory, then it is legitimate to ask whether the elements of alleged 'unrealism' in the experiment are parameters in the theory. If they are not parameters of the theory, then the criticism of 'unrealism' applies equally to the theory and the experiment. If there are field data to support the criticism, then of course it is important to [modify] the theory to include the phenomena in question, and this will affect the design of the relevant experiments" (Smith, 1982, p. 268).

First of all, such reasoning relies on the metaphysical thesis of nomic parallelism discussed in section 4.3. Recall the problem of identifying the 'structural' properties that determine the behaviour of economic (laboratory and non-laboratory) systems. Economic theory can give us a helping hand here: the theory identifies some explanatory variables as crucial and their interdependence is tested in the laboratory where no other factor intervenes. If the theory works there, we have no (theoretical) reason to believe that it should not work in the real world. If it does not work, we have no (theoretical) reason to believe that it should. The argument is directed against those who resist any (positive or negative) inference from the laboratory to the real world. If there are any reasons suggesting that the theory may be applicable in the real world while failing in the laboratory (or *vice versa*), then such reasons must be rendered explicit by modeling them into the theory. According to the metaphysics of nomic parallelism, an experimental result may be generalised to any situation in which the same *ceteris paribus* conditions hold: a theory failing in the laboratory and working in the real world (or *vice versa*) has ill-defined *ceteris paribus* conditions.[185] Smith's argument works by appealing to a precise methodological standard: let us call it the requirement of '*theory-completion*' (or simply 'completion', for short). As I shall argue in the next section, completion is strictly related to the requirement of generality or unrestrictedness imposed on scientific theories. I shall try to criticise both of them in section 4.5 below.

It should be noticed at this stage that the distinction between metaphysical and methodological (by completion, unrestrictedness, etc.) versions of parallelism is not a clear-cut one. What kind of rationale lies behind, e.g., the requirement of generality imposed on economic theories? If theories are supposed to capture the true form of social laws, does the generality requirement not after all depend on a factual matter concerning directly unobservable features of the social world (that is, on how economic laws really are)? In other words, the metaphysical claim that the laws governing social phenomena are indeed general in scope, unrestricted in applicability and cover all events (or that causally efficient properties are few in number and common to all individuals) seems to constitute an obvious rationale for the requirements of universality and generality imposed on economic theories,

[185] Some experimentalists go as far as to say that to shift the focus from real economies to theoretical models was the greatest innovation of experimental economics, and that an experiment "should be judged by the lessons it teaches about the theory and not by its similarity with what nature might have happened to have created" (Plott, 1991, p. 906).

and (*a fortiori*) for the methodological versions of parallelism [P3.1] and [P3.2]. We seem to be caught in a circular pattern of reasoning.

Let us then try to state a *purely* methodological parallelism thesis. A parallelism claim is truly methodological in character only if it is *not* a claim about how the social world is. It should rather start from the assumption that

> [P3] It is methodologically appropriate to act on the belief that parallelism holds.

Let us take [P3] as the underlying, general presupposition common to all methodological versions of parallelism. Acting as if parallelism held is useful if it is functional to achieving the goals of economic theorising or to fulfil the standards of science. In this section I have introduced some arguments which aim at justifying parallelism claims by appealing to standards such as unrestrictedness (generality) or completion. In the next section I shall discuss these requirements, to conclude that they stand on slippery ground.

4.5. A methodological detour

In order to criticise claims of methodological parallelism, it will be necessary to take a philosophical detour into the problem of how to interpret economic generalisations. The present section is self-sufficient and may be skipped without affecting the general argument. The conclusion will be that the incomplete, restricted generalisations of economics may not always be transformable into complete, general and unrestricted ones, and this undermines the requirements of completion and unrestrictedness.

Let us assume for the moment that scientific laws are appropriately represented as claims supporting counterfactuals in 'If...then...' form. The antecedents of such sentences represent the so-called 'initial conditions', a list of factors and circumstances causally relevant for the occurrence of the phenomenon of interest, which is described in the consequent. A nomic conditional can be either true or false, *tertium non datur*. With respect to a given particular situation or system, however, a law-like generalisation[186] is either (*a*) verified, (*b*) falsifed, or (*c*) vacuously instantiated - depending on whether (*a*) both the initial conditions

---

[186] I.e. a statement fulfilling all the requirements imposed on genuine laws except that of being true - a tentatively proposed candidate for a law of nature.

and the consequence are instantiated, (*b*) the initial conditions hold but the consequence does not, (*c*) the initial conditions are not satisfied. Were scientific generalisations formulated so as to include an explicit, complete list of factors and circumstances considered sufficient for the occurrence of the consequence, there would be no problem in discriminating between cases (*a*), (*b*) and (*c*). Unfortunately, scientists usually do not (because they *don't know* how to) exhaustively spell out the conditions for a certain relationship to hold. They rather behave in an 'opportunistic' manner, so that their nomic claims are better characterised as incomplete generalisations including *ceteris paribus* clauses in their antecedents, which are meant to capture unknown disturbances and 'background' conditions. These, together with the explicit initial conditions, are supposed to provide a set of jointly sufficient conditions for the consequent to hold, and thus at the same time give deductive closure to explanations and predictions from the generalisations.[187]

Given what just said, it is conceivable that some relationship (say, $y = f(x_1,...,x_n)$) working well in its intended domain be found not to hold in other contiguous domains (and vice-versa).[188] The fact that

(3) 'If it were the case that X_1*,...,X_n* then it would be the case that Y*',

'naively' formulated without *ceteris paribus* clause, is found to be false in a domain D_1, may simply mean that Y requires other (unspecified) circumstances above X_1,...X_n to hold for it to be instantiated. Thus, the modified claim

(3') 'In D_2, if it were the case that X_1*,...,X_n* then it would be the case that Y*'

may be true, if

(3'') '*Ceteris paribus*, if it were the case that X_1*,...,X_n* then it would be the case that Y*'

[187] All I have said so far can be rephrased in terms of other accounts of scientific theorising. In the next chapter I shall rely on the so-called 'Semantic View' of theories to state a different point. In the semantic framework, at any rate, the problem of specifying the sufficient conditions of applicability of a certain model arises in formulating the so-called 'theoretical hypothesis' (e.g. 'Every physical system is isomorphic to some model of classical mechanics' - see next chapter for further details).

[188] To simplify, one can identify the intended domain of economic theory with real world economies, and the contiguous domain with laboratory economies; but the intended domain may well overlap with the laboratory - it depends, of course, on what economists 'intend'.

is true, and the circumstances covered by the *ceteris paribus* clause are generally instantiated in D_2. It would not be wise, then, always to generalise from a failure in the laboratory to a failure in the field, in the absence of further evidence supporting this move. In a subsequent section I shall also provide an example of what the practical implications of this standpoint are, by focusing on a real controversial case of experimentation - the case of preference reversals already examined at length in the last chapter.

4.5.1 *Restricted, incomplete laws and growth-of-content methodologies*

The first rationale for methodological parallelism presented in section 4.4 was that only truly unrestricted (general) scientific theories should be deemed acceptable, and those failing in the laboratory clearly do not fulfil these requirements. The idea that genuine scientific laws are true generalisations of unlimited scope and applicability is part of some of the best philosophies of science of this century. It informs most reductionistic programmes, accounts of explanation such as Hempel's (1965) covering-law model, normative methodologies such as Popper's (1934/1959) falsificationism, and is part of the intuitive interpretation of our best scientific exemplars. Scientific theories do not have a domain of application written *explicitly* in their equations; yet, we know that classical mechanics and electromagnetic theory were supposed to hold for all physical bodies in the universe. Similarly, quantum mechanics aims at describing all particles irrespective of location and time and the same applies to general relativity for macro-objects. Of course, not all theories are 'theories of everything', but the latter have played for a long time the role of 'exemplars' of scientific excellence, which all other theories are supposed to imitate. The great successes of science have often been identified with a progressive explanatory unification of phenomena which had previously appeared heterogeneous (like electricity, magnetism and light) under the 'umbrella' of some theory (like the theory of elecromagnetism) of greater domain of application than each of its predecessors. Thus, 'growth of content' has been an important presupposition of philosophy of science: science progresses by reducing the number of theories and extending their content.[189]

[189]Cf. also Friedman (1974) and Kitcher (1981) for an illustration and defence of such a view in the context of the debate on 'scientific explanation'.

The so-called 'special sciences', however, seem to be at odds with the growth-of-content picture. A special science is a discipline concerned with the explanation of a certain specific realm of phenomena, usually at a non-fundamental level of analysis.[190] According to a popular picture of the hierarchy of the sciences, physics stands as the most fundamental discipline concerned with the explanation of the basic structure of the world. Chemistry, biology, psychology, and the social sciences are all branches of the tree, concerned with explaining phenomena supervening on the physical ones. The special sciences make use of nomic relationships that do not hold always and everywhere. General equilibrium theory does not describe how bargaining between two people ('pure bargaining', in economists' jargon) takes place, nor is the primary aim of consumers' theory to account for human behaviour in a mental hospital. Very few, if any, theories in the special sciences like economics can stand up to unlimited-scope standards.[191] In the best cases, they seem to describe some relationships holding in isolation from disturbances, when certain background conditions hold or, in other words, in a limited domain of application.[192]

This point can be formulated in terms of clauses and provisos capturing the influence of hidden factors and background conditions. Take a generic relationship such as

(4) $y = f(x_1, x_2)$.

By interpreting it as a *ceteris paribus* law, we are saying that there exist some so far hidden factors which allow the relationship to hold in a certain domain (i.e., where those factors are at work). We are saying, for instance, that

(4*) If $x_3 = a$, $x_4 = b$, *and other relevant conditions hold*, then $y = f(x_1, x_2)$.

[190] The present discussion relies mainly on Fodor's (1974) account.

[191] Of course there is a sense in which all scientific generalisations, when suitably formulated, have unlimited scope. Take a standard (and tedious) example as 'All ravens are black'. If reformulated as 'For all *x*, if *x* is a raven, then *x* is black', the generalisation has unlimited scope - i.e. it is a generalisation about all entities *x* (be them ravens or not). But of course the generalisation is 'vacuously verified' by all non-ravens. In other words, it is non explanatory with respect to all entities that do not feature all the properties (like 'being a raven') denoted by the predicates in the antecedent. In this - much more interesting - sense, 'All ravens are black' (like most generalisations in the special sciences) has a limited domain of application, and I shall stick to it in the rest of this chapter.

[192] Jim Woodward (unpublished) defends such a view in detail.

Notice that whereas some relevant factors (x_3 and x_4) and the values they have to take (a and b) can normally be specified, others cannot. (4*) is thus, first of all, an *incomplete* relationship, because it does not specify in the right-hand side of the equation the sufficient conditions for the values assumed by the dependent variable y; and secondly, a *restricted* one, because it is applicable - if at all - only where those (partly unspecified) conditions are instantiated.

Experimentalists eager to find support for their methodological parallelism claims could find in philosophers' writings a number of arguments in favour of the requirements of completion and unrestrictedness. Some arguments move from the difficulty to test (falsify or confirm) incomplete laws. A once influential interpretation of statements such as (4*) was that they are not even candidates for scientific laws. According to Popper (1934/1959), for instance, the falsifiability of a statement is a necessary condition for its being scientifically acceptable, and vague *ceteris paribus* clauses render laws unfalsifiable. This interpretation of *ceteris paribus* laws has been very popular among methodologists of economics, who have endorsed it in various forms. According to Terence Hutchison (1938), the non-falsifiability of *ceteris paribus* claims downgrades them to mere analytical statements. Klappholz and Agassi (1959) argue more cogently that they should be interpreted as synthetic but in principle untestable statements, and Blaug (1980) criticises incomplete generalisations precisely on these grounds. Such arguments are however too strictly related to the requirement of falsifiability, and therefore fail in the light of the well known fact that no theory taken in isolation is strictly speaking falsifiable, if we take falsifiability in Popper's logical sense.[193]

The standard view concerning incomplete and restricted generalisations is that they should be taken as 'promises' for a true, complete and unrestricted explanation. They have the status of "potentially explanatory theories", as Hempel (1965, p. 447) once put it. The validity of an incomplete generalisation can be established only as long as you are able to progressively incorporate the hidden background conditions as explicit variables in the right-hand side of the equation. Then, and only then, are there grounds to believe that you have hit a 'good' relationship, a promising starting point from which to reconstruct the whole puzzle. Izabella Nowakowa (1994) has an account of 'idealised' (i.e., in our terminology,

[193] To my knowledge, J. Alberto Coffa was the first one to clarify these issues in his doctoral dissertation (1973). For a detailed (and similar to Coffa's, but independently devised) account of the semantic and epistemic features of *ceteris paribus* laws, see Hausman (1989; 1992). The discussion on *ceteris paribus* and 'idealised' laws in physics has been revived in particular by Cartwright (1983). Hempel (1988) recognised some of the problems raised by the ubiquitous character of incomplete laws and vague clauses.

incomplete) laws which captures nicely such an intuition. You start with an unqualified generalisation like (4); in the light of anomalous evidence, you qualify it with something like (4*), partially completing it but restricting its domain of application; in an effort of generalisation you then find a new equation able to account for (4) and incorporating a new factor, say x_3, which becomes a truly explanatory variable; and so on and on.[194]

Notice that according to these accounts completion and the growth of empirical content are *necessary*, but surely not *sufficient* requirements imposed on new theories. Other restrictions may include the requirement that the theory be independently tested; that it predicts new facts and that these are corroborated at least now and then; that the new model is not obtained via 'ad-hoc' modification of the previous, falsified, one (whatever 'ad hoc' means - a highly debated question); that the new model is devised coherently with the 'spirit' of the research programme; that it is simple and elegant; that one sticks to economics' typical language and level of analysis, and so on. Growth of content is then just one among many requirements, but nonetheless a very important one (perhaps *the* most important one), especially in the tradition stemming from Popper's falsificationism.

Coherently with the Hempelian and the 'growth of content' tradition, a similar interpretation holds that *ceteris paribus* statements are to be accepted as genuine (albeit incomplete) laws only if they can be progressively expanded to incorporate the 'hidden' factors in the background.[195] A relation like (4*) is a genuine *ceteris paribus* law, in other words, if it is in principle derivable from a genuine general and unrestricted law by assigning particular values to some of its variables. For instance, in the four-variables case introduced above, $g(.)$ could be a general functional accounting for (4) and (4*) as special cases:

(5) $y = g(x_1, x_2, x_3, x_4)$.

Such a view surely captures one of the most fundamental heuristic strategies adopted by scientists when dealing with counterexamples and anomalies. Philosophers of science have discussed a closely related idea under the label of

---

[194]Cf. also McMullin (1985) and Laymon (1985) for a similar view on the role of 'successive approximations' in justifying idealised laws.

[195]Cf. Hausman (1989; 1992). Fodor (1991) calls it the 'completer' account of *ceteris paribus* laws: a *ceteris paribus* law is a true law if and only if there exist some hidden property which, if mentioned explicitely, could transform the *ceteris paribus* law into a true unrestricted generalisation.

the ‘correspondence principle’. According to the latter, a new theory should supersede an older one by explaining all its corroborated content plus some new facts the old theory could not account for.[196] Such an identification of the growth of knowledge with the growth of theories' content does indeed justify unrestrictedness and completion requirements such as those advocated by economic experimentalists. It is doubtful, however, whether we should accept the growth-of-content picture of science as foundations for the methodology of a special science like economics. As a matter of fact, the special sciences typically rely on relationships like (4) or, in the best cases, (4*), rarely being able to 'justify' them by means of anything like (5).

More precisely, the question is whether we should take the derivability of incomplete generalisations from deeper relationships as a *necessary* condition for the acceptance of theories - rather than merely as a heuristically useful rule. The view according to which ‘growth of content’ is necessary for the growth of scientific knowledge goes back at least to Popper and to the Standard View. Hempel and Oppenheim (1948), in their famous paper on the character of scientific laws refer to Popper as a source of inspiration,[197] and I shall therefore mainly focus on his account for expository purposes. The requirement of unlimited domain is in Popper’s philosophical framework an *ideal*. According to Popper, it is a necessary requirement of a superseding theory that its empirical content (the set of its potential falsifiers) exceeds the content of the superseded one. The content of a theory is a function of its *degree of universality* and its *degree of precision*. In particular, theoretical content increases with the level of universality: a theory or law *T’* is ‘more universal’ than *T* if it is logically true that the antecedent of the nomic conditional *T* implies the antecedent of *T’* (Popper, 1934/1959, § 36).[198] The scientist starting with a rather cautious hypothesis is

[196]Cf., e.g., Post (1971), Krajewski (1977), and Zahar (1983). Notice however that the correspondence principle is a very general precept: it says that a new theory supersedes by correspondence an older one if it is able to account for all the corroborated low-level empirical claims derivable from the latter, but it does not prescribe to keep the variables of the old theory in its equations. The new theory (unlike in Nowakowa’s account and in most accounts of *ceteris paribus* laws) may tell a completely new causal story. Moreover, we should appropriately speak of correspondence when one theory reduces to another in the limit: $f(x_1, x_2) = \lim_{x_3 \to a} g(x_1, x_2, x_3)$. The limit case is mathematically different from the case of a variable held fixed.

[197]Cf. Hempel and Oppenheim (1948, p. 266, footnote 26).

[198]Popper formalises the idea of degree of universality using relations of material implication (instead of counterfactual conditionals as in this chapter): suppose that T is $\forall x\,(Fx \to Gx)$, and T' is $\forall x\,(F'x \to Gx)$; then T' has a higher degree of universality than T if $\forall x\,(F'x \to Fx)$ holds. Moreover, T'' is more precise than T if T'' is $\forall x\,(Fx \to G'x)$, and $\forall x\,(Gx \to G'x)$ holds.

forced by Popperian methodology to formulate theories with greater domain of application, approximating genuine unrestricted generalisations in the long run.

The scientist, then, is *not* allowed to substitute an unrestricted generalisation with a generalisation of lower degree. Suppose we are dealing with a disturbing factor affecting the system under study, and that our best theories cannot explain it. We can distinguish two different strategies to take into account the impact of some disturbing factor on the theory's predictions: a *generalising* strategy and a '*domain'* one.[199] If we pursue the former, then we are looking for a new law covering the corroborated domain of the older theory (now 'explained' as a special case) *plus* the disturbing factor at hand. According to Popper (1934/1959) this is the right move: the scientist should look for a new theory explaining the incomplete, restricted law *and* the new factor. If we pursue a domain strategy, on the other hand, we simply circumscribe the domain of application of the theory by adding further specifications in the antecedent. Suppose *T* is 'For all *x*, if *Fx* then *Gx'*; a domain strategy will replace *T* with *T'*, the latter being 'For all *x*, if *Fx* and *Mx* then *Gx'*. According to Popper, such a move is an '*ad hoc*' one, because the refuted theory is replaced by a theory with a narrower domain of application. The new *T'* is a generalisation of lower degree, has a more limited domain of application than *T*, and a narrower class of potential falsifiers. It says what happens when an object has the property *F* and the property *M*, but does not pronounce on what happens when the object has the property *F* in general (with *and without M*). A new theory should be *content-increasing*: it should widen the domain of application of its predecessors.[200]

Such a view is still quite popular. Alexander Rosenberg, for instance, supports it vigorously in his recent book *Economics: Mathematical Politics or Science of Diminishing Returns?* (1992, ch. 5). Rosenberg proposes a methodological requirement of '*improvability*': a restricted and incomplete theory is acceptable only if it is improvable, that is, if it is possible to specify and add to it those (so far unspecified) factors that intervene to break the predicted

---

[199]The term 'domain strategy' is due to Musgrave (1981); cf. also Mongin (1988b).

[200] More about this, with examples, can be found in Popper (1957). Popper's prescription of looking for higher generalisations is the methodological, explicitly normative, counterpart of Hempel's interpretation of incomplete, *ceteris paribus*, idealised laws as statements deducible from more comprehensive and 'fundamental' nomic relations. Cf. e.g. Hempel (1965, pp. 83, 167, 345, 444 and 447). According to Herbert Feigl, "The aim of scientific explanation throughout the ages has been unification, that is, the comprehending of a maximum of facts and regularities in terms of a minimum of theoretical concepts and assumptions" (1970, p. 12). The identification of progress with 'continuous growth' and 'excess of empirical content' is characteristic also of Lakatos' 'Methodology of Scientific Research Programmes' (1970). See also Kitcher (1981) for other quotes and examples.

regularities at stake.[201] Since neoclassical economic theories do not seem to be very promising in terms of improvability, Rosenberg argues, they should be deemed unacceptable. Rosenberg's argument, in simplified form, proceeds in three steps. The first is a rejection of the behaviourist interpretation of rational choice theory,[202] a step which will not be discussed here. From this, Rosenberg argues to the necessity of endorsing a psychologistic interpretation of choice theory, where 'preferences', 'expectations', etc. refer to mental states of individual human beings. Finally, he presents Alan Nelson's (1986) 'psychoeconomics' as the best attempt to save microeconomics as an empirical *and improvable* scientific theory. Once the target has been set, he hits it by showing that 'psychoeconomics' is doomed to failure.

I shall not go into the details of the argument.[203] For our purposes, it is sufficient to notice that it is clearly built on a growth of content criterion: an incomplete law is acceptable only if transformable into a true unrestricted generalistion. Most economic practice, as Rosenberg argues, does not stand up to such standards. Reasoning of this sort, however, may in contrast lead us to think that completion or growth of content standards are inadequate to the special sciences, rather than the other way around. We shall see below that persuasive arguments can be put forward in support of a much weaker interpretation of such standards. The distinction suggested will be similar to Lakatos' (1970) distinction between methodology and heuristics, or to the Popper-Reichenbach distinction between the context of discovery and the context of justification. Whereas heuristic rules are context-dependent and relative to the problem-situation at

---

[201] The requirement of 'improvability' is taken from Hausman (1992), but with a twist. Hausman's position on the importance of growth-of-content requirements is ambiguous, and I do not have the space to comment on it here.

[202] The 'revealed preferences' interpretation according to which preferences are to be taken as patterns of observed choices.

[203] Here is a summary: Nelson's proposal is based on an analogy between economics and linguistics. Rational choice theory would stand to the theory of the ideal speaker as a more realistic theory of behaviour would stand to theories of actual linguistic performance. The role of the realistic theory is to provide an explanation of how actual behaviour emerges from the combination of the agent's ideal capacities with other factors which impede their full instantiation in the real world. Rosenberg sees two major problems in such an approach. First, "the list of [interfering] factors is so long, so heterogeneous, and so unsystematizable, that the competence/performance distinction is doomed to purely academic interest in the case of economics" (1992, p. 140). Secondly, beliefs and desires, as mental states in general, are intensional, the human mind being a holistic system. Factors "cannot simply be divided up and counted out. What we believe and desire is a whole not uniquely divisible into discrete parts. Thus the mind is inaccessible to a simple theory of rational choice competence" (1992, p. 147). The conclusions drawn by Rosenberg are not to the falsity of the psychological microfoundations of economic theory, but rather to the implausibility that an 'acceptable' scientific theory can ever be built on those shaky grounds (Rosenberg, 1992, p. 151).

hand, Methodology (capital 'M') is supposed to provide universally valid principles for appraising scientific practices and results.[204] When proposing their methodological interpretation of parallelism, I take it, experimentalists are putting forward principles of the latter kind. Only such universally valid principles would in fact be able to provide laboratory results with *a priori* validity. My proposal will be that parallelism claims be 'downgraded' to contingent principles, the validity of which is dependent on factual matters and sometimes even on the particular problem-situation at hands.

4.5.2 *Problems with completion and growth-of-content requirements*

Let us now see why the claim that the special sciences must provide complete theories with an unlimited domain of application should be rejected. Surely, more fundamental, 'deeper' laws encompassing earlier ones as special cases and extending the domain of application of the theory are better than less fundamental, 'shallow' ones. And surely a complete specification of all the sufficient conditions for the occurrence of an event is desirable. The point to be raised here, however, is whether we should assign to principles such as unrestrictedness and completion the status of methodological rules in the sense specified above. It is reasonable to look for increasing generality; but should we take it as a *necessary condition* for the growth of scientific knowledge?

A first reason to be sceptical of this proposal is pragmatic. As far as we know, the number of exceptions and background conditions may simply be too great for us to take them all into account (argument from the variety of the social world). Moreover, there are reasons to doubt that the deeper laws encompassing economics' restricted generalisations are expressible in the language of economics (argument from intrinsic linguistic limitations). One may have to shift to a different, much more complicated level of analysis in order to find deeper laws. In an often-quoted passage, Jerry Fodor makes such a feature a distinguishing characteristic of the special sciences:

> Exceptions to the generalisations of a special science are typically inexplicable from the point of view of (that is, in the vocabulary of) that science. That's one of the things that makes it a special science (Fodor, 1987, p. 6).

[204] For a clear statement of such a distinction, cf. Worrall (1988).

> Special science laws are unstrict not just de facto, but in principle. Specifically, they are characteristically '*heteronomic*': You can't convert them into strict laws by elaborating their antecedents. One reason why this is so is that special science laws typically fail in limiting conditions, or in conditions where the idealisations presupposed by the science aren't approximated; and, generally speaking, you have to go outside the vocabulary of the science to say what these conditions are (Fodor, 1989, p. 78).

To pursue a complete enumeration of the conditions captured by a *ceteris paribus* clause would force scientists to go outside their favoured realm of phenomena and level of analysis. The neoclassical economist for example would have to abandon his models of rational economic players and engage in a much deeper analysis of human psychology. The laws of psychology being *ceteris paribus* in character, in order to extend their domain one would have to move one further step down the ladder of microfoundations, to e.g. neurophysiology. Eventually, according to the ideal reductionistic picture, all the laws of the special sciences would find their justification in the generalisations of physics.

> [I]t may [...] be perfectly possible to explain the exceptions [of a special science] *in the vocabulary of some other science*. In the most familiar case, you go 'down' one or more levels and use the vocabulary of a more 'basic' science [...]. The availability of this strategy is one of the things that the hierarchical arrangement of our sciences buys for us (Fodor, 1987, p. 6).[205]

The feasibility of a reduction of economics to psychology and of the latter to physics cannot be ruled out in principle, but some empirical evidence must be put forward before we accept reductionism as the basis for our methodology. Special sciences such as economics exist precisely because it would not be practical to try explaining right from the start the phenomena falling in their intended domain in terms of more fundamental laws such as those of psychology or even physics (if they exist at all, of course). But if the criterion for scientificity is growth of content, then "the only real science is basic physics. For it simply isn't true that we can, even in principle, specify the conditions under which - say - geological

[205] See also Kincaid (1990) for arguments of these sort, with particular reference to the social sciences.

generalisations hold so long as we stick to the vocabulary of geology" (Fodor, 1987, p. 5).[206]

To conclude, let me recapitulate what just said. I have not tried to argue that it is wrong to look for a more general or deeper version of a theory suffering from anomalies - a version able to explain the confirmed consequences of the previous theory plus at least some of the anomalies. It would be great if we could do that all the time. The correct attitude is surely to look for more content when possible, and rely on other criteria in order to assess theories or research programmes that score badly in terms of growth. Neither I have claimed that to look for more general or deeper accounts is hopeless.[207] The position defended in this section is a much more pragmatic one: it starts from the recognition that the special sciences are concerned with generalisations at a non-fundamental level of analysis, and suggests that imposing any requirement (*necessary condition*) of continuous growth is not sensible. That would be equivalent to committing oneself to the reducibility of all sciences to a unique fundamental theory. To ask when and where a theory can be applied is a very important question, but in order to do this we may need help from other sciences. It is even more important if, as in the case of economics, the intended domain of the theory is very ambitious indeed: the set of all phenomena having to do with the pursue and exchange of goods.

I engaged in such a long argument because experimentalists often rely on considerations taken from 'growth-of-content' philosophies in order to justify

[206]Giving up the hope of completion raises automatically the problem of what economic generalisations say, after all. According to the standard view, they are 'promises' for a true, complete explanation of economic phenomena. One alternative solution is to argue, as John Stuart Mill and more recently Cartwright (1989) and Hausman (1992) do, that the independent variables in economic equations denote factors providing a *partial* causal explanation of how economic phenomena are brought about. Such causal factors act as 'tendencies' or 'capacities', providing a constant contribution to the production of the phenomenon denoted by the dependent variable across all circumstances - no matter what the background conditions are. The fact that in practice we might never be able to give a complete account of the 'other things' affecting the outcome does not affect the appeal of such a metaphysics *in principle*. Alternatively, one can argue (as Cartwright does in 1983 and in her more recent writings) that the models accurately describe what happens literally only when no other factors intervene.

[207]That would be the right conclusion if it was the case that the world was after all not governed by a small set of general principles at a very fundamental level. Arguments supporting this metaphysical view have been put forward by Patrick Suppes (1984), John Dupré (1993) and Nancy Cartwright (1994), who have argued that the universe is much more 'disordered' than scientists tend to assume. According to Cartwright's 'metaphysical nomological pluralism', for example, "nature is governed in different domains by different systems of laws not necessarily related to each other in any systematic or uniform way: by a patchwork of laws" (Cartwright, 1994, pp. 288-289). It is not my purpose here to discuss such a position. Pragmatic considerations such as those presented above suffice to reject the claims I was concerned with in the first place.

their parallelism claims. But if what said above is true, the standards advocated by these experimentalists would be not only irrealistic (pushing us to throw knowledge of the most useful sort in the bin) but misleading: growth of knowledge should not be necessarily equated with growth of theoretical content. The falsificationist parallelism standpoint [P3.1], for example, gives advice of the following sort: look for generalisations holding outside as well as inside the laboratory; test them in the laboratory; reject them if they are falsified there. But what if they *did* hold outside the laboratory (where they were supposed to hold in the first place)?

4.6. Empirical versions of parallelism (I)

Economists are surely keen both on seeing their discipline as an ‘inexact and separate science’, to use Dan Hausman’s (1992) terminology, and on subscribing to growth of content standards. Such commitments, however, pull them in opposite directions, and experimentalists do not constitute an exception. We have seen, for example, Charles Plott’s ideas on the importance of empirical counterexamples to theories, even when they come from laboratory experimentation, and how such an idea seems to be based on a generalist conception of theories. But Plott (1995) has also recently defended an interpretation of microeconomic theory called ‘the discovered preference hypothesis’, according to which rational behaviour emerges slowly at different stages in the development of an agent’s decision skills.[208] According to such an interpretation, rational choice models are consistent with data only when applied to the behaviour of experienced agents aware of the environment and able to anticipate the decisions of other rational choosers. A number of vaguely specified conditions must hold in order for the theory to be applicable: even if they have had a chance to practice with bargaining for a long time, a group of mentally ill people will hardly develop the capacity to anticipate the decisions of other rational agents.

Such examples call for an obvious comment: scientists are attached to certain specific ideas (something like a Lakatosian ‘hard core’), and often try to defend them at the expense of overall logical consistency. Thus, Plott ‘the experimentalist’ is ready to appeal to a requirement of theoretical unrestrictedness in order to defend the relevance of laboratory experimentation in general; Plott

[208]This idea can be found in a number of classics including Marshall, Pareto and Joan Robinson, and is therefore not particularly new. Ken Binmore (1999) has recently endorsed a very similar view.

'the neoclassical economist' uses every argument (including one involving a restricted interpretation of microeconomic theory) in order to save his favourite models from refutation. A discussion of such an opportunistic attitude would take us too far into the problem of evaluating scientists' reactions to anomalies, and their apparent conservativeness in particular. Such analysis has already been attempted by others,[209] and I have tackled the issue of the interpretation of economic theories only because it is relevant for some *prima facie* plausible arguments for parallelism. On this issue, Plott 'the neoclassical economist' (his deepest motivations notwithstanding, I should repeat) endorses in my view a more reasonable position than Plott 'the experimentalist'. Notice that the message of the 'discovered preferences' hypothesis is not a relieving one for the orthodox economist. This defensive strategy is *very* costly: the intended scope of economics would be drastically limited were we to take Plott's proposal seriously. Moreover, in the light of what said above about the special sciences, to interpret economic theory in a restrictive fashion opens the problem of how to explain and define the mechanisms that make the theory applicable in the first place. And this might be possible only with help from other neighbouring disciplines.

If a 'restricted' interpretation of economic relationships is legitimate,[210] the question shifts from *whether* human behaviour in general is consistent with, say, consumer theory, to *how much* real economic behaviour is of this sort, and how much is not. This leads us eventually to an *empirical version* of parallelism:

[P4] Parallelism must be checked empirically.

Such a minimal statement has the advantage of being free from problematic metaphysical commitments and from methodological standards that do not seem to fit economics particularly well. [P4] is applicable to nomic as well as to exhibited parallelism - both are, eventually, factual claims. It is thus compatible with a metaphysical thesis of high level of abstraction such as [P2.1] (Parallelism is exhibited whenever the same causally relevant conditions hold), without being in any way dependent on it. It has one great disadvantage, from the experimenters' point of view: it does not vindicate the interest of every experimental result *a priori*. The relevance of an experiment is dependent on its generalisability outside

---

[209] Cf. in particular, Hausman (1992) for the case of preference reversals, and Hausman and Mongin (1998) for a comparison of that case with another one.
[210] As it has been suggested above; but see Woodward (unpublished) for more arguments.

the laboratory, and this is a question which may be answered only *a posteriori* – when it can be answered at all.

Nevertheless, such an idea may sound a little paradoxical: did we not undertake experimentation in the first place precisely because dissatisfied with field evidence? The answer is yes, of course. To appreciate the *relative* advantages of controlled experimentation from the correct point of view, consider again the issue of preference reversals. Bisecting the problem (into a problem of internal validity and one of parallelism) proved in this case to be very useful. Experiments have supported the hypothesis that preference reversals occur (in the laboratory) to an extent that would have been impossible with field data. The parallelism step is more controversial, but it is surely better to take a firm small step plus another shaky small one, rather than one big shaky step at once. Moreover, in some cases the parallelism step may be rather uncontroversial, because based on commonly accepted assumptions. Indeed, the parallelism step is often invisible for this very reason: it is 'automatically' taken for granted. But (leaving the question of whether it is *justifiably* taken for granted aside) the fact that it cannot be seen does not imply that it is not there.

Experimenters have indeed put forward claims that can be read as versions of [P4], and that are quite incompatible with other, more radical, claims such as those reviewed in the previous sections. Vernon Smith for instance points out that

> When the theory performs well [in the lab] you [...] think, "Are there parallel results in naturally occurring field data?" You look for coherence across different data sets because theories are not specific to particular data sources. Such extensions are important because theories often make specific assumptions about information and institutions which can be controlled in the laboratory, but which may not accurately represent field data generating situations. Testing theories on the domain of their assumptions is sterile unless it is part of a research program concerned with extending the domain of application of theory to field environments (Smith, 1989, p. 152).

On this interpretation, parallelism is an empirical issue, to be settled on a case-by-case basis. *Parallelism is something to be achieved*, rather than an assumption we can rely upon. In some passages, Charles Plott similarly speaks of experiments as just one preliminary step in theory-appraisal:

> I do not think that experimental methods will replace field research. Economies found in the wild can only be understood by studying them in the wild. Field research is absolutely critical to such an understanding. However, the theories and models used in field research necessarily incorporate many judgments about assumptions, parameters and behavioral principles. The simple cases that can be studied in the laboratory can provide the data against which the importance of such judgments can be assessed. Economics is one of the few [*sic*] sciences that is fortunate to have both the field and the laboratory with which to work (Plott, 1991, p. 918).

The next issue to be discussed, then, is surely *how* parallelism questions may be answered empirically. For example, to what extent can one rely on laboratory evidence, and where do field data play a role? And what is the logical structure of the arguments put forward in order to justify parallelism claims? Discussing such questions will keep us busy for a while: the remaining part of this as well as the next chapter are indeed dedicated to different strategies for establishing parallelism. I shall now abandon abstract reasoning and turn to concrete examples. In the next section I shall start by analysing a real controversy that revolved around a parallelism problem - the by now familiar case of preference reversals.

4.7. Testing the robustness of preference reversals

Economic relationships are supposed to hold across a certain variety of situations, which define the domain of their application. The reasons why they hold, however, are not fully specified (and surely not formalised). Economists instead put forward a number of informal accounts of why one should be confident that some relationship holds where and when it is supposed to. As a typical example, take evolutionary and arbitrage 'stories': if an economic agent did not behave rationally at least most of the time, he would be exposed to exploitation and soon kicked out of the market. Or, similarly, whenever a market is not in equilibrium, there is the possibility of buying and trading so as to ensure a positive gain at no cost. The market's equilibrium would be restored by repeated arbitrage. To define these 'stories' as 'theoretical arguments' is somehow exaggerated, because they seldom take the form of rigorous theories, let alone an axiomatic form. (Using the convention adopted in the previous chapters, they are 'theories' with a small 't', rather than 'Theories', capital 'T'.) Still, they provide arguments for believing in the robustness of a relationship inside a certain domain, by pointing to some background circumstances that allow the relationship to hold.

When an experimental phenomenon is particularly indigestible, conservative scientists tend to insinuate that it cannot occur in the intended domain of the theory, and sometimes give reasons for this. Typically, they will say that there are some properties of real-world systems that prevent the anomalies from arising, and that such properties are not instantiated in the experimental environment where the anomalous phenomenon was generated. Notice that two empirical claims are made, and each has to be tested empirically. First, one has to try and reproduce in the lab the property at stake, so as to allegedly 'approximate' real-world conditions; this is the strategy followed by experimentalists working on the preference reversals phenomenon. If these extra conditions prevent the anomaly from occurring, one has to show that such properties are instantiated in the intended domain of the theory.

The research on preference reversals (PRs), for example, has broadly speaking gone through two main stages. As we saw in chapter three, a few years after Lichtenstein and Slovic's initial findings, economists started to devise experiments in order to test the robustness of reversals *in the laboratory*. They tried, in other words, to check whether the anomalous evidence (the observed choices) could be interpreted as an 'artefact' of the experimental techniques used in order to elicit agents' preferences. We have also seen that economists nowadays generally agree that preference reversals are a real laboratory phenomenon, rather than a mere illusion of our 'instruments' for the observation of preferences. Recently - perhaps frustrated by a series of failures to explain the anomaly away - some experimentalists have turned their attention to the robustness of reversals *outside* the laboratory. They are investigating whether preference reversals can be classified as 'artefacts$_2$', to use my previous terminology. Or, in other words, experimentalists are now concerned to establish or refute parallelism claims. The problem had been posed since the early days. In their influential paper, Grether and Plott argued that

> The key question is, of course, whether [PRs] should be of interest to economists. Specifically it seems necessary to answer the following:
> (1) Does the phenomenon exist in situations where economic theory is generally applied?

> (2) Can the phenomenon be explained by applying standard economic theory or some immediate variant thereof? (Grether and Plott, 1979, p. 624).

Berg, Dickhaut and O'Brien (1985) devised some experiments in order to test the robustness of the PR phenomenon to allegedly more 'realistic' conditions. The fundamental idea guiding their work is the one outlined above: economic generalisations should not be taken as exceptionless. They are on the contrary relationships that hold only where the 'right' conditions are instantiated. Thus, if we have a vague idea of what such circumstances may be, we can try to create an environment where they hold, and see whether they make any difference to the anomalous phenomenon (and, *a fortiori*, to the theory) in question.

> It remains an open question whether any mechanism, particularly one which would exist in situations where economic theory is generally applied, can substantially reduce or alter [PR] inconsistencies. The mechanism considered in our work is an arbitrage procedure. In general, the possibility of arbitrage in a market setting leads to the conclusion that there cannot be market inconsistencies such as two prices for the same commodity (Berg, Dickhaut and O'Brien, 1985, p. 33).

The background mechanism responsible for the relationships of consumer theory to hold is therefore illustrated as follows:

> If preference reversals exist in an exchange setting, they create an arbitrage opportunity. A subject having been arbitraged is expected to realize that inconsistencies will be exploited and therefore to reduce both the rate and size of reversals (Berg, Dickhaut and O'Brien, 1985, p. 34).

Notice that the background mechanism is not formally modeled nor incorporated into the theory of consumer's behaviour. This is not necessary in order to test the theory's relevance for the PR phenomenon. Experimentalists are only interested in checking whether *when* the arbitrage mechanism is at work, learning takes place and reversals tend to disappear. The theory could then be applied as it is. Experimenters in this case are not looking for a theoretical *explanation* of the mechanism, which is therefore kept confined in the 'background' of the theory.

The design of the Berg-Dickhaut-O'Brien experiment is far from trivial. One problem with these experiments is to combine the exchange task needed to money-pump 'inconsistent' subjects with the already rather complex machinery of a standard PR experiment. The BDM procedure, for example, can be used to determine the *real* reservation price in the exchange mechanism only if subjects are assumed to be constantly and absolutely risk-averse; otherwise, their buying and selling prices will differ (cf. Berg, Dickhaut and O'Brien, 1985, pp. 34-35). Another incentive procedure especially invented by O'Brien to solve this problem was therefore used in the experiment. After the announced prices were elicited, the subjects were required to trade with the experimenter on the basis of their announced prices and choices.

The design also controlled for another variable, i.e. repetition. As with Plott's 'discovered preferences' hypothesis, the central idea is that a period of learning may be necessary to acquire the decision skills conjectured by models of rational choice such as expected utility theory. By controlling for arbitrage and repetition, Berg, Dickhaut and O'Brien discovered that PRs do not disappear under these conditions. The frequency of reversals was even slightly higher when subjects experienced arbitrage than when they did not, thus replicating a similar effect already observed in benchmark experiments such as Grether and Plott's (1979). Their dollar magnitude was, however, substantially decreased (from a mean value of 4.10 to 2.52 dollars). Repetition definitely diminished both the frequency and the dollar magnitude of reversals - the number of reversals per subject dropping from 36% in the first trial to 27% in the second; and the value from a mean dollar magnitude of 4.02 to 2.58. Not surprisingly, the two effects combine with each other, so that the most significative reductions can be observed in groups subject to *both* repetition and arbitrage. Berg and his collegues eventually recognised that the phenomenon was eroded but 'did not go away'.

The results of this experiment did not discourage the sceptics. Chu and Chu (1990) a few years later devised a variant of this experiment that controlled for the effects of *complexity* and of *repeated arbitrage*. They simplified the standard PRs design and exposed their subjects to a longer series of money-pumps. Whereas simplification alone did not seem to have great impact on the phenomenon, repeated arbitrage did. Moreover, the effects of learning were observed to be persistent: once exposed to repeated money-pumping, subjects acted more consistently with standard economic theory in immediately subsequent tasks.

Other experiments were inspired by similar ideas. Knez and Smith (1987) tried to design an experiment featuring both the characteristics of typical PR experiments and of market experiments known to deliver results consistent with standard economic theory. The crucial idea behind PRs is that people adopt different 'response modes' for different tasks, depending on whether they are engaged in pricing or choosing. Knez and Smith assigned to their subjects a fixed amount of money and bets to be traded under double-auction rules. Their hypothetical valuations and choices for each bet were elicited, and used to plot demand-supply curves for the market. Then the subjects were asked to trade their bets. Inconsistencies between valuations, choices and trading behaviour were observed, but the frequency of reversals was reduced substantially (from 60 to 40%) by the trading experience.[211]

All these experiments, to sum up, showed that preference reversals *can* be reduced in certain situations. Two questions, however, arise: the first concerns *how much* arbitrage and repetition (for example) is needed in order to eliminate PRs or at least reduce them to negligible proportions; and secondly, are these preventatives active in 'real world' markets?

4.8. Empirical versions of parallelism (II)

What can the PRs case teach us about the problem of empirical parallelism? When some anomalous phenomenon has been produced in the laboratory, economists are often in the uncomfortable situation of deciding about its relevance without any empirical evidence that suggests that it has some real-world counterpart.[212] The question then is: Should we take the experimental result

[211] As Knez and Smith rightly point out (1987, p. 132-133), the results of such experiments are not easily interpreted, because the disappearing of inconsistencies and the convergence to market prices do not univoquely confirm standard individual choice theory. The theory of market efficiency does not require that demand behaviour be consistent with individual expected utility theory: there may be convergence to clearing prices even though agents' *real* preferences are inconsistent. The two theories, in this sense, are quite independent. The problem of inferring from 'market experiments' to the (non-)robustness of PRs is therefore severe, and affects also other more recent research. Cox and Grether (1996) have tried to elicit preferences by using three different devices: a standard Becker-DeGroot-Marschak procedure, a sealed second-price auction, and an English clock auction. The sealed-bid auction is a typical pricing procedure, whereas the English clock auction and the Becker-DeGroot-Marschak mechanism are choice procedures. The effects of repetition were particulary strong in the two auctions, where reversals diminished substantially and asymmetries disappeared. A correlation between new and old prices generated in the markets was also observed. It is not clear, however, whether such 'anchoring effect' is due to a convergence of agents' revealed preferences towards their 'true' valuations or to simple imitation.

[212] Assuming, of course, that 'casual' observations of, e.g., people's inconsistencies in everyday life do not count as scientifically relevant evidence.

seriously, or just shelve it until some evidence of transferrability has been presented? The debate between enthusiasts for and sceptics about experimentation usually revolves around this question. For example, suppose a phenomenon (*O*) inconsistent with a widely accepted proposition of economic theory (*T*) is observed in the laboratory. The sceptics usually insinuate that *O* is observed only when certain background conditions *B* are instantiated, that conditions *B* were indeed instantiated in the laboratory, and that in 'real-world' economic systems circumstances *B*' (which are incompatible with *B* and thus prevent *O* from occurring) are the rule. The theory, then, is saved by adding (in an informal way, usually) *B*' as an auxiliary assumption. Notice however that the overall argument involves not only the (conjectural) prediction that *O* in fact does not occur when both the antecedent in *T* and the additional *B'* are instantiated, but also the factual claim that the preventative *B' is* indeed instantiated in the real economies the theory is supposed to apply to in the first place - two empirical claims.

We can now identify two different rules of method. According to the *Strong Version of Empirical Parallelism*,

> [P4.1] The burden of proof lies with the economist willing to save *T*, who has to show that
> (i) *T* obtains for some B_i;
> *and*
> (ii) B_i is instantiated in the real-world economies whose behaviour the theory is supposed to explain.

In contrast, the *Weak Version of Empirical Parallelism* asserts that

> [P4.2] The burden of proof lies with the experimentalist willing to reject *T*, who has to show that
> *either*
> (i') *T* obtains for no B_i;
> *or*,
> (ii') none of the B_i for which *T* obtains is instantiated in the real-world economies whose behaviour the theory is supposed to explain.

The auxiliary assumptions B_i are supposed to be borrowed from some stock of reasonable or scientifically relevant propositions. Where does the burden of proof lie? Let us start with (i) and (i'): to know the exact conditions for the applicability of a theory, its boundaries and limitations, is surely crucial for an efficient use of the theory itself. Nobody will deny that it is important to know, for example, that a theory previously believed to be of unlimited domain of application has rather to be interpreted as a restricted generalisation. (Although, it is worth repeating, to require that all the background conditions and mechanisms necessary for the theoretical propositions to hold be incorporated in the theory's formalism is unreasonable.) Thus, were a theory supposed to hold unconditionally, in case of falsification the burden of proof would lie with the defender of the theory to point to some background conditions that would save the theory from refutation (case (i) above). It would in fact be impossible for the challenger to show that there do not exist *any* B_i which would save the theory; holding [P4.2](i') would be a too dogmatic and conservative attitude.

The subsequent step consists in testing whether it is really the case that when the conditions specified in T plus the appropriately chosen B_i are instantiated, the phenomenon O does not occur. Such a test can sometimes be carried through in the laboratory. All the preference reversal experiments seen above test the relevance of some background conditions taken to be important by common economic wisdom, although not made explicit in the formal theory of consumer's behaviour. Given the incomplete character of economic generalisations, such experiments are indispensable in order to infer from the observation of an anomaly in the laboratory to the falsification[213] of a theory unable to account for it. Testing the robustness of a phenomenon to certain background conditions is however only half of the job, and needs to be completed by a third step.

Once the appropriate background conditions have been identified (if they have been at all), there still remains the question whether - as implied by (ii) - the defender needs to show that these conditions really are instantiated in the real-world cases of interest, or whether the challenger is required to show that they are not (as prescribed by (ii')). Subscribing to (ii) would amount to claiming that *Parallelism is supposed to hold until some empirical reason to the contrary has been given*. Holding (ii'), on the contrary, would imply that *Parallelism is not supposed to hold until some empirical reason to the contrary has been given*. Notice that by framing the question this way one is implicitly assuming that the empirical test of parallelism advocated in [P4] cannot be exclusively

[213]Of course I am here and elsewhere in this chapter speaking of methodological falsifications (as opposed to logical ones) in the sense of Lakatos (1970): the *decision* to abandon a theory in the light of the available evidence *plus* some extra-logical methodological criterion.

experimental in character. Parallelism always involves a correspondence between two kinds of phenomena, one in the laboratory and the other in the real world. Establishing parallelism requires the collaboration between laboratory and field evidence. (In the next chapters I shall show, by means of new examples, how this can be done in practice.)

Such an abstract analysis can now be put at work and applied to the case study examined in this chapter. The categories above may in fact be used as a sort of ideal-typical characterisation of a parallelism problem and how it has to be tackled. The first thing to notice is that economists engaged in the preference reversals case did not go very far into investigating parallelism. Chu and Chu, for example, summarise their results by saying that the PR phenomenon appears to be vulnerable to a 'marketlike' environment (1990, p. 910), but such a claim is at least ambiguous. As Colin Camerer has put it,

> These results suggest that in an environment where preference reversal is a recognizable, costly mistake that outsiders can spot and exploit, then people can learn to switch their expressed preferences (or reduce the size of any discrepancy). But there is no evidence of whether subjects who are disciplined this way then learn to express preference more consistently in the future, or whether reversals actually persist in natural settings (Camerer, 1995, p. 661).

Chu and Chu seem to suggest that PRs tend not to occur (or occur only very occasionally) in real markets, whereas their experiments prove at best only that PRs tend not to occur in those markets where repeated arbitrage is possible, and where subjects have the possibility of learning to be consistent. They may prove, in other words, that the orthodox theory holds where the 'right' background mechanisms are at work. *Whether* these are (always? sometimes?) at work in the intended domain of the theory, is a different question, which should be settled on an empirical basis. Only half an argument against parallelism, then, has been carried through. Experimentalists sceptical about the PR phenomenon have tried to 'mimick the real world', but their conception of what is 'real' was shaped by their (*a priori*) views about the structure of market economies. Informal but powerful arguments such as 'if there were inconsistencies in the market, then someone would exploit them until they disappear; therefore there cannot be inconsistencies in the market' play a crucial role in justifying the invisible parallelism step from the laboratory to the real world.[214] The parallelism step is in this case (partly) supported by economists' preconceptions, rather than by empirical proof. One cannot fail to notice, in the debates on controversial experimental findings, a tendency to replace empirical statements about factual matters - such as required by the principles [P4.1] and [P4.2] discussed above - with general statements which lack solid empirical foundations. These arguments stand on one leg only, and fall short of the ideal standards I have proposed above. A philosophical analysis such as the one attempted above may indeed help economists to improve their methodology, to go beyond rhetorical arguments and to concentrate on factual matters.

To sum up, in this section two abstract versions of empirical parallelism have been proposed and discussed. Both stem from the assumption that parallelism claims must be checked empirically, but differ in the weight they assign to anomalous evidence. The strong version prescribes taking the theory as falsified until it has been shown that the target, 'real-world', economies customarily feature background mechanisms or conditions preventing the anomaly to occur. The weak version of empirical parallelism asserts that the theory should not be rejected until it has been shown either that there exist no background mechanisms preventing the anomaly from occurring, or that such mechanisms are not at work in the target systems at stake. I have suggested that the weak version is too dogmatic, and that the 'aggressive' attitude implied by the strong version sounds more reasonable. In the preference reversals case, this suggests that the discovery of the phenomenon has shifted the burden of proof towards those willing to save choice theory from refutation. In fact, experimentalists have been looking (and partly successfully so) for preventatives of the reversals phenomenon. Finally, I have argued that these defensive strategies are not effective until it has been shown that background mechanisms such as learning from arbitrage are also at work in the 'real economies' whose functioning economic theory aims to explain.

4.9. Conclusion

The preference reversals case seems to substantiate my analysis in sections 4.1-4.6 above. Whether one should revise an economic theory in order to account for anomalous experimental evidence is a matter that cannot be settled *a priori*. No

[214] Notice, by the way, that such reasoning is based on the assumption that economic agents behave as the rational greedy maximisers postulated by standard economic models. But preference reversal experiments are meant to test (and to question) precisely the empirical validity of these models!

experimental evidence can falsify a theory unless parallelism with real-world economic phenomena has been established; but neither can a phenomenon be disregarded on the basis of purely *a priori* reasoning. Whenever it is possible to refer to some background conditions that make the relationship robust across the range of situations it is supposed to be applied to, such conditions should be included in the experimental environment in order to test the anomaly's relevance. It is unreasonable however to require that *all* the background conditions, provisos, and *ceteris paribus* clauses be made explicit before a theory is deemed acceptable. The requirement of completion upon which experimentalists' arguments for parallelism are based should therefore be rejected. How 'deep' one is supposed to dig into the realm of background conditions is a context-dependent matter. I have argued that the reasons why parallelism should not hold must be made explicit and testable, but not necessarily incorporated into the theory's formalism. Finally, I have shown that the word 'empirically' in statement [P4] ("Parallelism must be checked empirically") cannot mean 'experimentally *only*'. The empirical version of parallelism put forward in the previous section must therefore be interpreted in this sense:

> [P4*] Parallelism must be checked by collaboration between experimental and field data.

The form of such collaboration has to be further investigated, and I shall try to say something more precise about it in the following chapters. To conclude and summarise this discussion of experimentalists' arguments: *a priori* reasoning will not convince sceptics of the relevance of laboratory experimentation to economics. Such arguments may have played a useful rhetorical role to convince editors and referees to accept experimental papers, but stand on slippery grounds. Too much should be thrown away were we to interpret them literally. I propose, then, to take the issue of parallelism the following way: the results achieved in the laboratory are not *in principle* relevant nor irrelevant to assess economic theories - their (ir-) relevance must be somehow justified *a posteriori*.[215] Indeed, experimentalists have often done this; given that the arguments one finds in their methodological papers are mainly 'aprioristic' in character, it is necessary to search for genuine parallelism arguments somewhere else. In the next chapters I shall look in their scientific papers for the *empirical* procedures they have devised in order to support inferences to parallelism in specific cases.

[215] For a similar conclusion drawn from different considerations, see also Starmer (1999).

One final conceptual and terminological distinction. I suggested earlier that the issue of 'parallelism' in experimental economics is very close to the issue of 'external validity' discussed by other social scientists, in particular by psychologists. If the two were identical, it would be wise to simplify and speak of external validity all the time. There is, however, one good reason to keep the two separate. The problem of external validity is often discussed by scientists as that of *generalising* a certain experimental result outside the specific circumstances of the experiment (e.g. to generalise beyond the particular population, the specific instruments used, place and time, etc.): from a claim of the form 'In laboratory L_i, with subjects S_i, selected according to criteria C_i, trained according to method T_i, pretested in the way P_i, etc., if X then Y' to a claim of the form 'For all subjects in all circumstances, if X then Y'. By shifting from the former to the latter statement, one is assuming (and this assumption may of course be based on very good empirical reasons) that the 'background experimental conditions' $B_i = \{L_i, S_i, C_i, T_i, P_i,...\}$ do not matter for the relationship between X and Y to hold. To put it in a different way, when external validity in this sense has been established one is entitled to claim that 'X causes Y *unconditionally*'.[216]

One can already see what is wrong with this: linking external validity to unconditionality is excessively demanding, and surely goes much beyond what economic experimentalists should (and often *do*) demand. If 'if X then Y' is true only in population A or circumstances B, then unconditionality requires that what distinguishes A or B be spelled out and included among the X's. To endorse this principle is equivalent to commit oneself to a complete, explicit specification of *all* the relevant factors, or to make do without *ceteris paribus* clauses in nomic and causal claims. This, as I have argued at length, is an unreasonable requirement. If we think that at least some of the generalisations of the special sciences capture genuinely causal or nomic relationships, then we better allow for a relative invariance of such relationships (relative to the 'right' background), as proposed for instance by Jim Woodward (1997 and unpublished). Speaking of parallelism

[216] The idea is that for all X and Y, *X causes Y* iff $P(Y|X) > P(Y|{\sim}X)$ *for all background conditions B*. See Humphreys (1989, pp. 72 and following) for a detailed presentation of the principle and a philosophical discussion. Humphreys (1989, p. 79, footnote 19) points explicitly to the relation between unconditionality requirements and external validity referring to Campbell and Stanley's (1966) classic on quasi-experimental methodologies. Unconditionality is more or less closely related to a number of other requirements proposed by statisticians and philosophers in order to tell genuine causal claims from spurious probabilistic correlations, including: randomness criteria (a number of statisticians since Fisher), 'maximal specificity' (Hempel), 'objective homogeneity' (Salmon), 'resiliency' (Skyrms), 'relevance invariance' (Cartwright) and 'contextual unanimity' (Dupré).

instead of external validity can help us to mark such a distinction. To achieve parallelism is not equivalent to achieve unconditionality: it is perhaps better characterised as a weak form of external validity limited to some typical variations in the background conditions. It is to establish that a certain experimental result can be 'exported' from a set of (experimental) systems L_i to a set of ('real-world') systems W_i. It is still a *local* inference, because economists do not have to endorse the view that the relationships they subscribe to must hold across *all* kinds of situations. Of course parallelism claims are ampliative, but do not require absolute generality. In order to capture this distinction also at the level of terminology, it is therefore useful to keep using the word 'parallelism'.

Chapter 5

Experiments as Mediators
Testing the 'winner's curse' hypothesis

> 'We have seen that there are three sorts of bed. The first exists in nature, and we would say, I suppose, that it was made by god. No one else could have made it, could they?'
> 'I think not.'
> 'The second is made by the carpenter.'
> 'Yes.'
> 'And the third by the painter?'
> 'Granted.'
> 'So painter, carpenter and god are each responsible for one kind of bed.'
> 'Yes.'
> (Plato, *The Republic*, Part 10, 597b)

5.1. Introduction

The case of the last chapter - that parallelism is an empirical hypothesis, rather than a claim to be endorsed or dismissed on *a priori* grounds - of course leaves open the question of how can parallelism be established. I have already discussed an example in the last chapter, but the case of preference reversals turned out to be partially disappointing - experimentalists went only half way to showing that the phenomenon could not occur in the real world of untamed economic phenomena. In this chapter another example will be presented and discussed, in the attempt to lay bare the structure of parallelism arguments.

Taking parallelism seriously will, as I shall try to show, force us to replace the standard, rather simplistic view of the logic of experimenting with a more complicated account closer to real scientific practice. According to the standard hypothetico-deductive model of scientific testing, some prediction about an observable event (*O*) is deduced from a theory (*T*), some initial conditions (*IC*), and auxiliary assumptions (*AA*): (T & IC & AA) $\Rightarrow$ *O*. If the predicted event is observed, the whole body of theory, initial conditions and auxiliaries is confirmed; otherwise, it is falsified and has to be revised. The hypothetico-deductive model does not place any requirements on where *O* comes from: it may be a casual observation in a natural setting as well as a highly artificial phenomenon produced in a constrained environment. One of the appealing features of the hypothetico-deductive model is indeed its wide scope of application. By looking deeper at the details of testing in different sciences, however, some important differences will emerge.

To begin with, we should introduce a distinction based on the kind of empirical data typically used in a given discipline, and thus speak of 'laboratory' vs. 'non-laboratory' sciences. According to Ian Hacking,

> Laboratory sciences are those whose claims to truth answer primarily to work done in the laboratory. They study phenomena that seldom or ever occur in a pure state before people have brought them under surveillance. Exaggerating a little, I say that the phenomena under study are created in the laboratory. The laboratory sciences use apparatus in isolation to interfere with the course of that aspect of nature that is under study, the end in view being an increase in knowledge, understanding, and control of a general or generalizable sort (1992, p. 33).

Non-laboratory sciences, then, are those whose claims to truth do *not* answer primarily to work done in the laboratory. More importantly, they can*not* answer primarily to work done in the laboratory, because the primary aim of such sciences is to explain and control *non*-laboratory phenomena.

I should stress that the laboratory vs. non-laboratory distinction may be more a historical and contingent than an essential one. Physics was once a non-laboratory science, becoming a primarily experimental discipline in the 17th Century. But the laboratory revolution was long and gradual. Two centuries later physicists were still allowed to worry about the realistic nature of their theories and experiments. According to Galison and Assmus (1989), indeed, in the Victorian age a 'morphological' school challenged the 'analytical' approach to scientific investigation that has since won the day. Whereas the 'analysts' searched for fundamental laws by tearing nature apart in their models (by abstraction) and in their laboratories (by shielding and manipulating), the morphologists looked with scepticism at such practices and rather tried to recreate full natural phenomena in the laboratory by means of 'mimetic' experiments. The morphologists, in other words, cared about parallelism: what can the highly constrained and disciplined environments created in the laboratories teach us about the unconstrained, complex outside world?

Nowadays worries of this kind are not very common. Physics has gone through two laboratory revolutions: from an Aristotelian science concerned with the explanation of spontaneously occurring phenomena by means of unaided observations; to a Galilean science investigating natural phenomena in the 'ideal' conditions of the laboratory; to a science, finally, whose main questions and answers are generated in the laboratory.[217] Sciences like economics or medicine, in contrast, supplement controlled experiments with other methods of enquiry. Non-laboratory sciences, then, use experimentation, but in a different way than laboratory sciences do.

In this chapter, I shall argue that experiments in non-laboratory sciences are just an intermediate step on the ladder leading to scientific knowledge 'of a general or generalisable sort'. I shall rely on some examples of experimental work done in economics, and try to show that experiments in sciences like economics play the role of '*epistemic mediators*'. They help to bridge the gap between a theory and its target domain of application, but not in the straightforward way imagined by the proponents of the hypothetico-deductive model. Experiments are just one part of a rather complicated engine for testing scientific theories. The role of experiments will be explicated by analogy: I shall try to show that experiments are used in many respects like models, relying on a specific view of modelling recently put forward by R.I.G. Hughes (the DDI account), Margaret Morrison and Mary Morgan (models as 'mediators').

The problem will be examined from two distinct points of view. First, I shall focus on the nature of laboratory experiments in economics. Secondly, I shall turn to epistemology and try to show what role experiments play in the process of confirming a scientific hypothesis. Despite their being conceptually distinct, the two aspects of the problem are clearly connected. What role experiments can

---

[217] One may wonder whether this shift is a progressive one: surely to focus primarily on what can be studied in the lab inmplis a drastic restriction in the domain of the discipline. But, surely, this shift has payed off in terms of technological applications. I discuss the use of experiments for the creation of new economic 'technology' in Guala (unpublished).

play depends of course on what the experiments are, and conversely, it is mainly by looking at how experiments are used that we can tell what they are.

The main case study - Kagel and Levin's experiments on the winner's curse phenomenon - will be introduced in section 5.2. I shall argue that two separate issues arise concerning any laboratory experiment in economics, namely: Can a given phenomenon be produced in the laboratory? And: does that phenomenon also occur in certain non-laboratory situations of interest? These questions will be treated separately in section 5.3. In section 5.4 I shall illustrate in what sense models may act as mediators. In section 5.5 the notion of 'demonstrating' using models and experiments will be introduced. In section 5.6 the case study of winner's curse experiments will be examined again in order to show how an inference from the laboratory to the outside world may be rationally formulated. Then (section 5.7), I shall try to show that such an inference is based on a combination of analogical and inductive reasoning.

5.2. (Re)producing the winner's curse phenomenon

If you care about the problem of parallelism, you are bound to ask two independent questions: Do theories provide adequate explanations of what is going on in experiments? And secondly, Do experiments correctly reproduce real-world situations, properties, or phenomena? In order to discuss these issues with a concrete case in mind, it is useful to select an example of experimentation where the two are kept neatly separate. The case of experiments on the 'winner's curse' phenomenon is a good example from this respect, and throughout the chapter I shall go back and forth from methodological analysis to the case study. In this section I shall just introduce the example and show how the first question above ('internal' validity) was addressed by experimental economists. Then, in sections 5.5 and 5.6, I shall come back to the winner's curse, focusing on parallelism.

In 1971 the Atlantic Richfield Company claimed that the constantly low profits derived from the exploitation of oil leases in the Gulf of Mexico was the result of their being the victims of a 'winner's curse' (Capen, Clapp and Campbell, 1971).[218] Oil leases are auctioned by a federal agency, the Outer Continental Shelf (OCS). Auctions of this kind are, technically speaking, 'common value auctions' - auctions in which the value of the auctioned item is the same for all

[218]For an introductory survey of the literature on the winner's curse phenomenon, see Thaler (1988).

participants, but initially unknown to all. A crucial part of the bidding game, then, consists in trying to estimate the true value of the lease. When the participants fail in this estimation, the winning bid is likely to turn out overoptimistic and the exploitation of the lease not profitable.

The claims of the Atlantic Richfield Company were suspect: they had an interest in convincing other companies to be more cautious in their valuations, and their move may have been a disguised invitation to act as a cartel by bidding less on the licences. On the other hand, a winner's curse phenomenon may have *really* been hidden below the data. How can we decide? The problem is that field data do not help very much to settle the dispute - since they are not able to convey information about crucial variables such as agents' private valuations or the real profitability of an oil lease in the long run.

John Kagel and Dan Levin (1986) tried to give an answer by reproducing the winner's curse phenomenon in the laboratory. In their experiment, information was provided to the bidders about the possible value of the item to be auctioned by communicating to each individual agent a value x_i drawn from a uniform distribution $[x_0 - \varepsilon, x_0 + \varepsilon]$, where x_0 is the real value of the item (i.e. the sum experimenters will pay the winning bidder) randomly drawn from a uniform distribution on an interval $[x^*, x^{**}]$. Experimenters communicated to their subjects the range of ε, and computed for them the upper and lower bound for the value of x_0.[219] The format of the experiment was dictated by auction theory, which models bidders' uncertainty as a random draw from a lottery of the above kind.

A game-theoretic account of auction mechanisms has been available since the sixties thanks to the pioneering work of William Vickrey (1961). Vickrey devised a model known as the 'independent private values model', where each bidder is supposed to be aware exactly of the value of the auctioned item, but does not know the value to other bidders. Such an assumption seems to be satisfied in auctions of, e.g., antiques, that will be privately enjoyed by buyers who do not intend to resell them. Oil leases do not seem to be that kind of good: their value, as we have said, is unknown but approximately the same for all bidders. Wilson (1977), and then Milgrom and Weber (1982), proposed a generalised theory of auctions able to account for the private-value and the common-value models as special cases. The auction is modeled as a non-cooperative game played by

[219]Min $\{x_i + \varepsilon, x^{**}\}$ and max $\{x_i - \varepsilon, x^*\}$, respectively.

expected-utility maximising bidders. The players are assumed to adopt equilibrium strategies - in the standard sense of a Bayes-Nash equilibrium in which, given everyone else's strategy, no agent can do better than he is presently doing by changing his strategy. The common value model is based on four basic assumptions: (1) That values are common and unknown to all; (2) that the bidders are symmetric; (3) that the pay-off is a function of bids alone; and (4) that bidders are risk-neutral. The former three assumptions appear empirically justified in the OCS case; risk-neutrality is needed for analytical reasons (assuming, e.g., risk-aversion leads to ambiguous results in a number of cases).[220]

The solution of the standard bidding model is known as 'non-cooperative equilibrium with risk-neutral bidders' (or RNNE for short), and predicts that the agent with the highest private signal (denoted x_1) will generally win the auction. If bidders are rational maximisers, as the RNNE models assumes, the one with the highest x_i is supposed to revise the expected value of the item in the light of the fact that his private information signal is the highest. In technical terms, the need for this revision is known as the 'adverse-selection problem'. Denoting the expected value conditional on having the highest information signal as $E[\, x_0|X_i = x_1]$, a winner's curse occurs every time the actual estimate of value exceeds the latter, i.e. whenever

(WC) $E[x_0|x_i] > E[\, x_0|X_i = x_1]$.

In this case, in fact, the winner fails to take into account the adverse-selection problem, and so will experience on average negative profits. The inequality above is better characterised as 'the winner's curse *hypothesis*': unlike the RNNE model, it does not provide a full explanation of the bidding process. It is defined as a contrast case: it conjectures that real bidders are not fully rational and fail to revise their expected values correctly. If the RNNE model is right, and bidders really are rational maximisers, the winner's curse should not occur, and evidence such as that presented by the Atlantic Richfield Company should be explained in a different way. The aim of the experiment devised by Kagel and Levin was to show how the data may result from a particular mechanism by reproducing it in the laboratory. The experiment had a very precise, predetermined target.

[220]For an introduction to auction theory, cf. Milgrom (1989); for a more comprehensive survey, see McAfee and McMillan (1987).

Kagel and Levin (1986) constructed their argument for the existence of the winner's curse by controlling for the number of subjects and public information. To begin with, (i) they ran experiments with a 'large' number of bidders (6-7) and experiments with a 'small' number (3-4). When the number of competitors is large, a rational maximiser is supposed to take into account two opposite considerations: one should bid more aggressively because the signal values are more congested, but less aggressively because the adverse selection problem becomes more severe. A RNNE bid function taking into account these considerations requires the bids to remain constant or to decrease when there is a growing number of competitors.[221] If the winner's curse explanation is right, by contrast, higher bids should be observed as the number of competitors increases. Varying the number of bidders thus provides a means of discriminating between the two rival hypotheses.

(ii) Some experiments involved only private information signals, whereas others involved public information: bidders were asked to provide a first evaluation under knowledge of x_i only, and then a second one after having been given some additional public information signal x_p (the lowest of the private signals formerly distributed, x_L, is particularly convenient for analytical reasons). The public information control is useful to provide insights into the bidding mechanism. In RNNE, in fact, public information is supposed to raise the bids of all the subjects who have not had the highest private signal; this should put pressure on the x_1 bidder (the winner, according to RNNE) and therefore diminish his profits by almost one half.[222]

Kagel and Levin (1986) observed two results: (i) in 'small group' experiments, the winners bought the items at a profitable price, but the profits were considerably lower than those predicted by the RNNE model (65.1% of the latter). In 'large group' experiments, the winners experienced losses on average. (ii) In auctions with a small number of bidders, the injection of public information raises prices; when the number of bidders is large, in contrast, prices fall contrary to the RNNE prediction. Both results are consistent with a winner's curse explanation. Winners, ex hypothesis, overestimate values; public information tends

[221] See Kagel and Levin (1986) for the quantitative analysis behind such a hypothesis.

[222] From $E[\Pi|W] = 2\varepsilon / (N + 1) - Y$ (where N is the number of bidders in the auction and Y is a negative exponential becoming rapidly negligible as the value of x_i departs from extremely low values), to $E[\Pi|W, X_L] = \varepsilon / (N + 1)$. See Kagel and Levin (1986) for the details of such a prediction.

to reduce uncertainty about the true value of the item, so that bidders with the highest private information can revise their evaluations.

Kagel and Levin, thus, tried first to run experiments that could teach us something about the functioning of laboratory common value auctions. I shall return to their experiments later (sections 5.5-5.6), to see how they were also designed to support an inference from the 'internal validity' of the winner's curse explanation to 'parallelism'. Before then, some more philosophical weaponry must be introduced.

5.3. Parallelism and underdetermination

Kagel and Levin demonstrated that a winner's curse phenomenon can be created in a laboratory economy - but what about the 'real world', the target phenomenon that originally motivated their investigations? Is a winner's curse interpretation of the OCS data legitimate? A further step is needed in order to claim that the same phenomenon observed in the laboratory lies hidden behind real-world empirical data. Two well known and distinct methodological problems arise with any laboratory experiment: the Duhem-Quine problem and the problem of causal underdetermination.

The *Duhem-Quine problem*, as we have seen in chapter three, consists in the impossibility of logically determining the inadequacy of a given theory on the basis of evidence and deductive logic alone. When faced with a falsifying observation, there always is the logical possibility of revising a peripheral assumption so as to save a given theoretical claim. The scientist, in other words, is never logically compelled to blame any particular component of a cluster of theories. This problem is also related to the so-called 'underdetermination of theories by data' thesis: logically speaking, a potentially infinite number of theories can account for any body of evidence (no matter how great). Which is just another way of saying that positive inferences from data to theory can only be inductive in character. As argued in chapter three, these problems cannot be solved, but can at least be reduced in the laboratory. In practice, in fact, in any historically given controversy only a finite number of competing theories exist, and one can discriminate between them by means of 'quasi-crucial' tests such as those devised by Safra, Segal and Spivak, Starmer and Sugden, and others in the preference reversals case, or Kagel and Levin in the case we are presently concerned with.

Experimental testing can help to reduce the problem of underdetermination of theory by data and thus confirm that a certain explanation is able to account for laboratory evidence. But it cannot eliminate nor even reduce the problem of *causal underdetermination*: different causal processes may generate similar patterns of data in different situations. In order to generalise a laboratory result, a further step has to be made: one has to show that the system constructed in the laboratory is the same as the one at work in the real world (the 'target', from now on), and that the similarity between artificial results and real phenomena is not illusory. Economists - as I have shown at length in the last chapter - have named it the problem of 'parallelism': what does the behaviour of laboratory economic systems tell us about phenomena observed in other, sometimes more complicated or very different, situations?

My proposed answer will be that experiments act as 'mediators'. The idea of 'mediating entities' has been already used to characterise the notion and role of scientific models.[223] I shall build on this idea and speak of 'epistemic' mediators: experiments, according to my account, constitute an intermediate step in a more general procedure aimed at supporting a given theoretical explanation of field data. Experiments are just one among many kinds of 'mediating entities', as I shall try to argue below. In the next section the notion of models as mediators will be presented and briefly discussed. Then I shall turn to experiments, and try to show that a common denominator of all mediating entities is their being simpler, more manageable, and more controllable systems which can be manipulated with the aim of understanding the functioning of a complex, little manageable, and partially or totally uncontrollable system.

5.4. Models

Let us start with the notion of model. There are several kinds of scientific models, and different taxonomies have been put forward in order to classify them.[224] Instead of engaging in a review of the literature, however, I shall focus on one particular kind of models, which have been called, according to their function, 'mediating models'.

---

[223]Cf. Morgan and Morrison (1999) for an illustration of such a view.
[224]Cf. e.g. Giere (1979; 1988) and Redhead (1980).

Let me first introduce an example - a famous economic model, put forward by Thomas Schelling in his *Micromotives and Macrobehavior* (1978). Schelling's goal was to give an account of racial segregation in American cities, but this was achieved by describing an imaginary checquerboard with a particular tessellation, upon which coins of two kinds (dimes and pennies) move according to specified rules. He shows that when some specific conditions hold, in particular when certain preferences about the occupants of each coin's neighbour squares hold, then certain dispositions of the coins on the checquerboard will follow. Informally (but the rules of the game are given by Schelling in game-theoretic form), each coin moves in the attempt to escape from areas where an overwhelming majority of coins of the other type prevails (say, 2/3 or more). Every time a coin is surrounded by a majority of another type, it is moved. *Via* successive reshuffling, it is shown that a complete separation of dimes and pennies may be produced on the checquerboard.

Schelling's paper is an exercise in analysing the dynamics of a toy-model, but the story about the model-world is intended to support a hypothetical explanation of the evolution of racial segregation in the real world.[225] In a recent paper devoted to discussing Schelling's model, Robert Sugden argues that "moving from the model to the hypothesis required a step in the argument which most readers would be willing to make, but for which no formal justification was available" (forthcoming, p. 12). According to Sugden, such a step is an *inductive* one.

> What Schelling has done is to construct a set of *imaginary* cities, whose working we can easily understand. In these cities, racial segregation evolves only if people have preferences about the racial mix of their neighbours, but strong segregation evolves even if those preferences are quite mild. In these imaginary cities, we also find that the spatial boundaries between the races tend to move over time, while segregation is preserved. We are invited to make the inductive inference that similar causal processes apply in real multi-ethnic cities. We now look at such cities. Here too we find strong spatial segregation between ethnic groups, and here too we find that the boundaries between groups move over time. Since the same effects are found in both real and imaginary cities, it is at least credible to suppose that

[225] The 'dictionary' translating the model into a real-world representation is as follows: coins = people; dimes and pennies = two races; areas = neighbourhoods; separation = racial segregation; rules = people's preferences; etc. As Dan Hausman pointed out to me, Schelling's contribution can also be read as the exploration of an interesting and counterintuitive *possibility*, rather than as an explanation of real world behaviour. Sugden (forthcoming) discusses and dismisses this interpretation in section 3 of his paper.

> the same causes are responsible. Thus, we have been given some reason to think that segregation in real cities is caused by preferences for segregation, and that the extent of segregation is invariant to changes in the strength of such preferences (Sugden, forthcoming, p. 34).

Let us notice, first of all, that the activity of modelling in economics has to do with the construction, and the theoretical description, of 'model-worlds'. Such systems are mostly abstract entities, existing only in the minds of those who happen to read, for instance, Schelling's book. But in principle a real, material chequerboard could be manufactured with its dimes and pennies, and a 'segregation game' played for real (Schelling actually invites the readers to do so). Schelling's game-theoretic account of segregation is *true* of the checquerboard described in his book - and trivially so, because the checquerboard system is designed so as to satisfy Schelling's game-theoretic formalism exactly. I have chosen Schelling's theory of segregation because the model there (the checquerboard, the dimes and pennies, and the rules of the game) is easily visualisable and separable from the theoretical formalism. But the same applies to all theories (including, e.g., the theory of auction we have started with): provided that their axioms and principles are consistent, an abstract model is identified of which the formalism is true.

It should be easy to grasp this idea for those trained in mainstream economics. Take for example the Walrasian auctioneer that is central in general equilibrium models. It is a typical model in the sense above: it is an abstract entity, because no real market uses tâtonnement to determine prices (although a few market institutions are similar to the Walrasian auctioneer). And it is an entity of which the theory's equations are true: if such an institution existed, then Walrasian equilibrium theory would fit it perfectly. Indeed, Walras in the fourth edition of the *Elements of Pure Economics* seems to suggest that the term 'tâtonnement' refers to the technique of solving a system of symultaneous equations by iteration.[226] The ambiguity (tâtonnement as what the auctioneer does, or as what the theorist does?) in this case just helps us to understand the nature of models in a better way.

We have here a first, clear sense in which models can be said to act as 'mediators': a theory's formalism (i.e. the set of 'axioms', 'principles' or 'laws'

[226]On the Walrasian auctioneer and its various possible interpretations, see De Vroey (1998).

of the theory) is not applied straight to reality, but is first and foremost asserted to be true of an ideal model-world, and *then* is suggested to be somehow relevant for the understanding some real-world phenomenon. The path from theory to the real world is broken down into at least two steps. Some philosophers of science have proposed, indeed, to take the model-worlds of which theoretical principles are true as the primary unit of analysis defining what a scientific theory is. Such an idea - put forward among others by Patrick Suppes (1969), Wolfgang Stegmüller (1979), Ronald Giere (1979/1997; 1988), Bas van Fraassen (1980), Frederick Suppe (1989) - is usually referred to as the 'semantic view' of theories. According to the semantic view, a theory is mainly defined by a family of models, rather than by any specific linguistic assertion about them, or about the real world.[227]

Philosophers have learnt from logicians and meta-mathematicians to think of models as *objects*. More precisely, as *structures* (sets of objects with their properties and relations) of which some axiomatic system is true. Scientists, and economists in particular, nevertheless often refer to syntactical entities (sometimes even to purely syntactical, uninterpreted, ones) as 'models'.[228] But here it is useful to stick to a well-defined and rigorous terminology, and speak of models the way semantic theorists do. I shall follow Ronald Giere's (1979/1997; 1988) account, and say that a scientific theory is made of two elements: a family of models, and a 'theoretical hypothesis' asserting that the models stand in some particular relationship with the real world. For instance, that segregation in American cities evolves in a somehow 'similar' way as dimes and pennies move on Schelling's checquerboard. Or that the Walrasian auctioneer provides a good caricature of how market forces operate.

The Semantic View was put forward with some precise goals in mind. First of all, by focusing on semantic aspects rather than on the syntactical aspects of a theory, semantic theorists intended to turn away from problems with language which had obsessed the logical positivists. A number of alternative axiomatic systems can define the very set of structures for which a theory holds - or in other

[227]The semantic view of theories is in reality less monolothic than I am pretending it to be here. To be precise, one should for instance distinguish between the 'semantic', the 'structuralist', and the 'predicate' approach; others like to distinguish the 'set-theoretic' from the 'state-space' and the 'representational' approach. These distinctions are too fine-grained for my present purposes, but see Hausman (1992, ch. 5) for a discussion of which version fits economic modelling practice better.

[228]On the differences and similarities between mathematicians' and applied scientists' use of the term model, cf. Suppes (1969).

words which are *semantic models* of that theory. To claim that a theory is a set of models (rather than a set of axioms) amounts to say that it is relatively unimportant which syntactic formulation of the theory has been chosen. What really matters are the models. These will include structures like abstract set-theoretical entities, systems of dimensionless mass-points and ideal harmonic oscillators (for classical mechanics), or checquerboards inhabited by dimes and pennies (for Schelling's theory of segregation). From an empirical point of view, the fact that a theory's equations are true of such structures is not very informative, though. Clearly, it is not scientists' primary goal to describe such structures. The latter are useful simply as an intermediate step on the way to explain or describe something else. The reason why scientists focus on models rather than real systems is pragmatic: models are simpler, more manageable, and easier to describe than reality.

A theory's axiomatics usually includes 'fictional assumptions' that do not have any real-world counterpart. These can be of various sorts. Some of them, like the assumption that economic agents are rational greedy maximisers, are false when taken as unrestricted generalisations, but may be true of most situations in the intended domain of the theory. Others, like the homogeneity condition in consumers' demand, or the macroeconomic assumption of zero interest rates, are almost always false of the real-world situations the theory is supposed to explain, but may conceivably hold in some other (unlikely) circumstances. Sometimes they are necessary for analytical reasons, sometimes they are just heuristic assumptions for the sake of illustrating or developing the model at hand. Finally, there are assumptions which cannot possibly be true of any entities in the real world, like the assumption that prices may be irrational numbers, or that the economy has an infinite number of agents. Some (if not most) systems of axioms or laws, then, have semantic counterparts (models) only in some mathematical heaven, or in a Platonic world of ideas. The second aspect that makes the semantic view appealing, then, has to do with truth: one can concede that scientific laws are true of models only, and argue that some other relation holds between the models and the real world. (Of course this does not exclude in principle that in some special cases there are real systems among the models of the theory.[229]) Such a relationship has been characterised in a number of different ways, from van

[229] In a similar vein, economists sometimes ask whether a certain economy is, e.g., a 'Keynesian economy'. It is better to resist such talk for two reasons: first, because of the problems with idealisation discussed in the main text above, suggesting that real-world models are rare; secondly, because a distinctive characteristics of models is their being simpler and more manageable than real-world systems, and conflating the two does not help us in the analysis of the role and nature of models.

Fraassen's (1980) 'isomorphism' to Giere's (1988) 'similarity' and Hughes' (1997) 'denotation', and there is a lot of discussion about which one is best. I shall not be concerned with this problem here, however. Let us just keep in mind that theoretical laws (axioms, principles) are true of models, and that a distinct hypothesis is often needed to link the latter to the real world.

To sum up then, and most importantly for our purposes, the Semantic View accounts for the fact that the axioms and laws of a theory are usually about some intermediate entities. The latter usually have a number of 'nice' features, which make them preferable to the target systems falling in the intended domain of the theory. This not only helps to solve some puzzles (like the highly idealised character of most scientific theories) which are not easily dealt with by the traditional framework; it also provides us with a powerful idea that may be extended beyond the realm of theoretical models to explicate the role of other mediating entities like experiments, material models, and simulations.

5.5. Mediators

As anticipated, I shall try to explicate the nature and role of economic experiments by analogy, arguing that they are used in many respects like models. The idea of 'mediating models' has been at the centre of a research project on 'Modelling in Physics and Economics' carried out at the London School of Economics, and has generated a number of case studies.[230] In this section I shall just summarise some of the main tenets of the 'mediators' approach and show how they can be used for my purposes.

5.5.1 *Demonstrations*

Mediating entities have three main characteristics: they are *partly independent* both from high theory and from the systems they are supposed to explain; they '*stand for*' some real-world systems of interest (which I shall call 'target systems'); and they can be *manipulated* in order to learn something about the real world. I shall start with the latter feature. R.I.G. Hughes (1997) has called it the capacity to 'demonstrate', and in this section I shall rely on his account - the 'Denotation, Demonstration, Interpretation' (DDI) account of models - in order to extend it to experiments in the non-laboratory sciences.

[230] Cf. Morgan and Morrison (eds. 1999), Morrison (1998b), as well as the symposium on mediating models in *Philosophy of Science*, Vol. 64, Supplement (Proceedings of *PSA 1996*).

Models display the important property of having internal mechanisms, which determine their evolution under certain conditions. This is a property of both theoretical (mathematical) and material models.

> Like an analogical representation, [a mathematical representation] presents us with a secondary subject that has, so to speak, a life of its own. In other words, the representation has an internal dynamic whose effects we can examine. From the behavior of the model we can draw hypothetical conclusions about the world over and above the data we started with (Hughes, 1997, p. S331).

Hughes calls the process of producing certain consequences by manipulating the model, '*demonstration*'.[231] The word has a long history and is particularly well chosen: it was common practice in the 17th Century to 'demonstrate' in public with experiments and models, and of course its connotation overlaps also with that of mathematical proof.[232] The basic idea implicit in demonstrating is that of triggering a mechanism and observing what its consequences (e.g. a theorem, or a physical effect) are. As an example of a material model, Hughes takes a model of light waves as water waves in a ripple tank. One 'demonstrates' the propagation of light waves by producing a process of propagation of water waves.

> Whereas in the 17th century geometrical theorems were said to be 'demonstrated', nowadays we demonstrate physical phenomena in the laboratory. Mathematical models enable us to demonstrate results in the first sense, material models in the second (Hughes, 1997, p. S332).

To sum up: models can be theoretical (abstract) or material (concrete) in character. Both have the capacity to 'demonstrate'. When demonstrating with a theoretical model, one usually explores the properties of an abstract structure and the consequences of some changes in it. When demonstrating with material models, one takes a concrete system that is supposed to display at a certain (more or less abstract) level of analysis a behaviour analogous to that of the target system one is interested in. The material system can then be used to mimick the

---

[231]The idea of manipulability as a crucial characteristic of mediating model is central in Morgan and Morrison's (1999) account. Hughes' own version should be seen in the context of that (i.e. the 'Models in Physics and Economics') project.

[232]Steve Shapin (1988) analyses the social role of public 'demonstrations' in detail.

behaviour of the target system. Hughes' notion of 'demonstration' nicely captures this activity of 'playing with' entities which are not necessarily the set of sentences defining the syntactical component of a scientific theory, nor the target systems the theory is ultimately supposed to be applied to. Morgan and Morrison (1999) have called such entities 'mediators'.

By moving from theoretical to material models we get 'closer and closer' to experimental systems (I shall try to make this intuition more precise later, especially in chapter six). The latter too may be seen as mediating entities, useful for their capacity to demonstrate. Both material models and experimental systems feature concrete, material mechanisms which the scientist can use in order to understand the functioning of a target system. In the laboratory sciences experimenters 'play' with the target system itself; in the non-laboratory sciences it is sometimes possible to manipulate the target system in a non-laboratory environment, but this is more often difficult, costly, dangerous, even impossible, and the inferences drawn from uncontrolled experiments are hardly reliable anyway.

We can see this also outside economics: one of the principal justifications for animal experimentation in medicine is that direct research on humans (the 'target systems') are not only morally unacceptable, but also methodologically problematic. Confounding factors of various sorts make the collection of an unbiased sample of data particularly difficult. While investigating the influence of electromagnetic fields on cancer rates, for example, North Carolina researchers tried to select a sample of subjects by picking up telephone numbers randomly. But apparently lower-income people are less likely to be at home during the day. The data suggested that brain cancer is highly correlated with exposure to electromagnetic fields, but also with traffic density, maternal smoking, and breast-feeding - which are all correlated with poverty. Animals, as opposed to humans, seem to be more controllable, and to provide data free from biases of the above kind.[233]

5.5.2 *Representativeness*

[233] Arguments of this sort can be found in official documents of the American Medical Association. The specific example is taken from LaFollette and Shanks (1996, ch. 2), who provide a most interesting discussion of the methodological problems (especially 'parallelism') facing animal experimentation.

The word 'demostration' is apt also because it evokes the fact that the experiment is neither the target system nor a representation of it - it just *stands for* it. In this sense, the experimental auction of our example does a similar job. One may say that it 'represents' the real OCS auction, but in the sense of being a surrogate for it (as a mediating entity, it is a 'representative', rather than a 'representation'[234]). Thus, there must be some relations linking the mediating object to the target system. More precisely, we need one function taking us from the target to the mediator, and another inverse one to take us back from the mediator to the target after the latter has been used for investigative purposes.

Hughes (1997) puts it as follows: a theory aims at providing an explanation of the behaviour of a certain system; in order to do so, it points to some model which will *denote* the system. *Interpretation* is the inverse of denotation: it is a function taking us from the model to the real system represented by the model. Usually, the process of interpretation takes place after some non-trivial consequences (predictions) have been drawn from the model. Denotation is the first step, interpretation the last one; in between, a demonstration takes place. The detour through models, from and back to the target system can, according to Hughes, be represented as in figure 1.

[234] For such a distinction, see Hughes (1999).

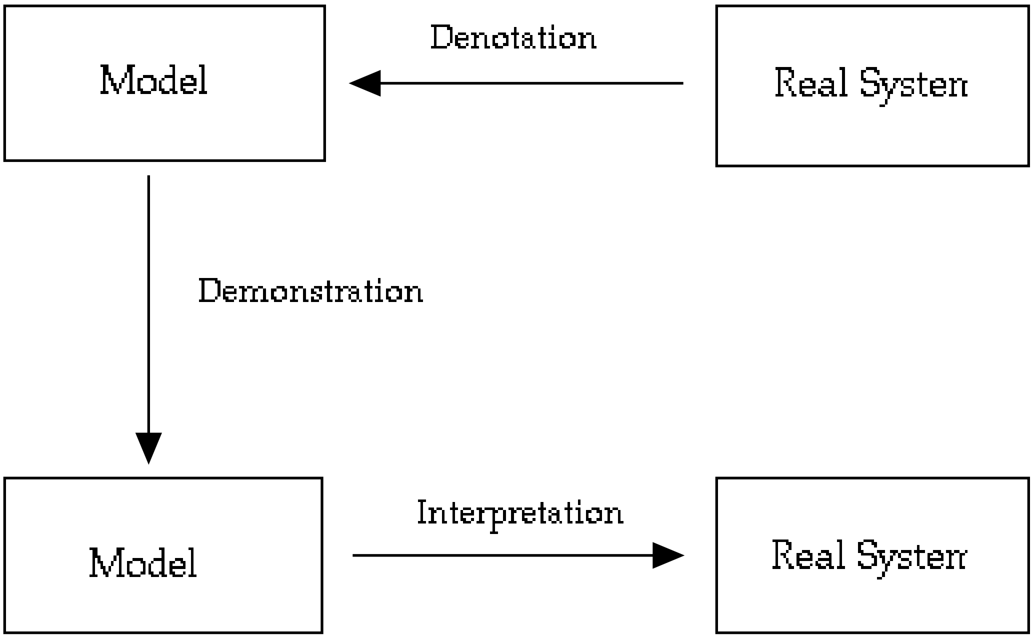

Figure 1: Hughes' DDI account

As I have said, the relationship between models and the systems falling in the theory's intended domain of application has been characterised in a number of ways ('similarity' and 'isomorphism' being two well-known alternatives to Hughes' 'denotation-interpretation' account). The debate is philosophically sophisticated and it is not the purpose of this chapter to contribute to it. In the remaining sections I shall use the DDI account with its terminology for reasons of simplicity, but I have to stress that the 'demonstration' component is the crucial one as far as I am concerned.

Thus, *laboratory experiments in the non-laboratory sciences demonstrate with experimental systems that 'stand for' the target systems of interest*. Such is my main claim concerning the nature of experiments in economics. Any mediating entity is linked either to the real world or to another mediator by means of some particular relationship. According to Giere (1988), for example, theoretical models are linked to some real-world entity by a 'theoretical hypothesis', stating what relationship holds between the model and the real world. In the case of experiments, similarly, a '*parallelism hypothesis*' has to be put forward saying that the laboratory system stands in some particular relationship to the target. Kagel and Levin, for instance, claim that their laboratory auctions are common value auctions like the OCS ones. (They are partially isomorphic with respect to auction theory, in the sense that they share all the observable properties modeled in the theory's language.) The two steps leading from a given economic model to the real-world phenomenon it intends to explain can therefore be represented as in figure 2.

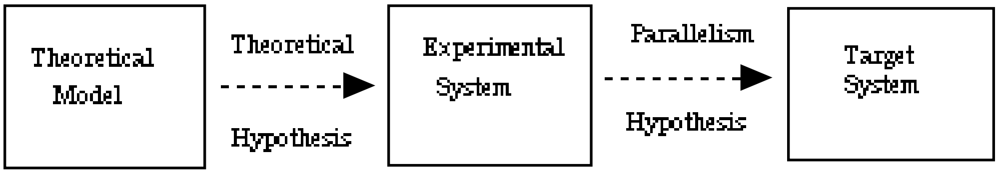

Figure 2: The path from theoretical models to the real world

In the case examined in section 2, there were two competing models, the winner's curse and the RNNE one. I shall devote the next sections to discuss how parallelism hypotheses are tested by reviewing the OCS case in detail: What kind of data are needed to confirm or refute a parallelism hypothesis? And how are they used? Before then, it is worth discussing the third and last crucial feature of mediating entities, namely their 'independence'.

5.5.3 *Independence*

The know-how used in building mediating models comes from different sources. Scientific models are rarely entirely theoretical or empirical in character. They are usually hybrid objects, and for this reason they can function like tools or instruments. "It is precisely because models are partially independent of both theories and the world, that they have this autonomous component and so can be used as instruments of exploration in both domains" (Morgan and Morrison, 1999, p. 1).[235]

Experiments too are autonomous from theory and the systems they are intended to represent. They are obviously autonomous from theory from an ontic point of view, but are also partially autonomous from theory from an epistemic viewpoint: a lot more than theoretical knowledge is needed in order to build, run, and interpret an experiment correctly. Experiments are designed only in part according to theoretical contraints. The path from the target system to models and experiments can be represented as in Figure 3. Let us discuss the diagram with the example of Kagel and Levin's winner's curse experiments in mind.

[235] Again, for a number of examples supporting this claim, see the case studies in Morgan and Morrison (eds. 1999).

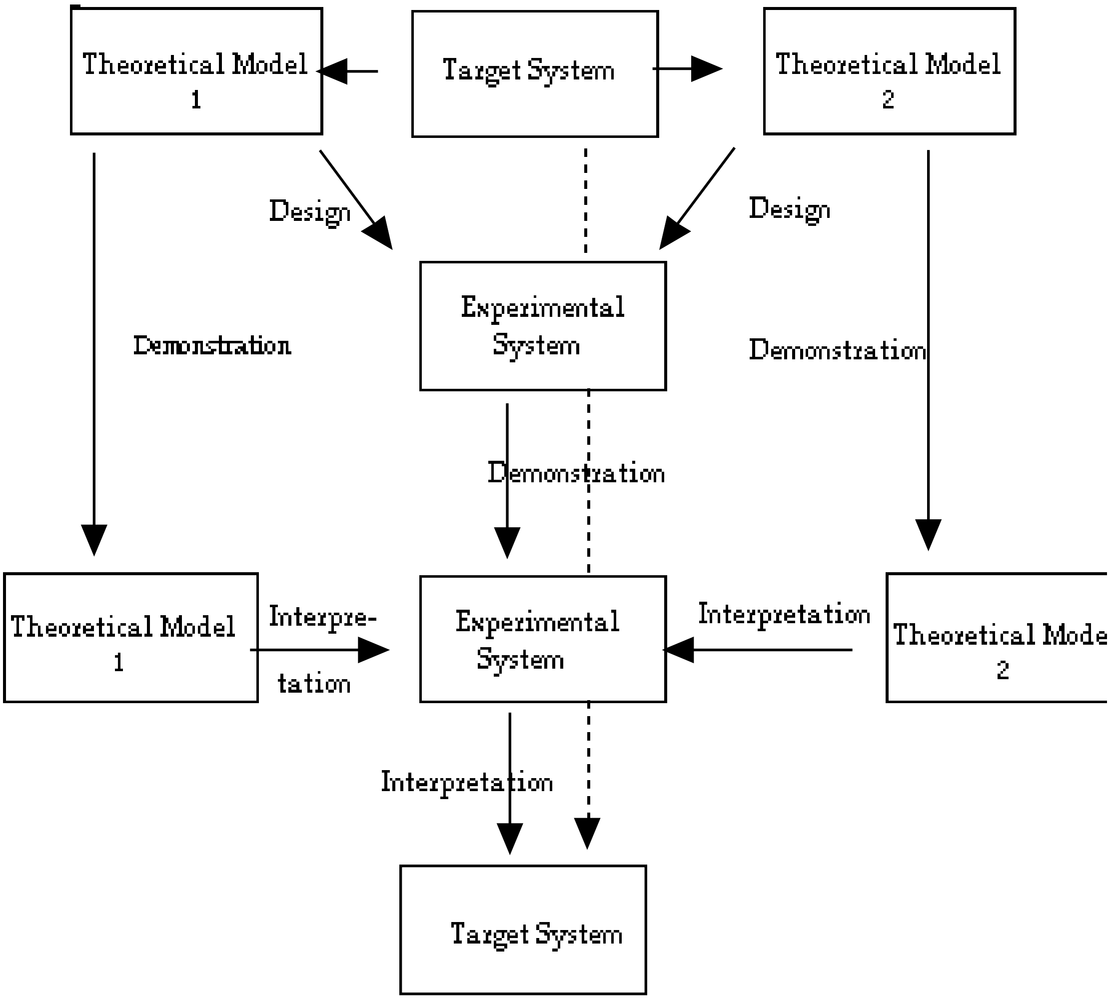

Figure 3: Experiments as mediators

The target systems to be represented consist of the OCS auctions, and oil companies' low profits is the phenomenon to be explained. The first move is to choose a theoretical model providing an explanation of this phenomenon. There are in our case two rival candidates, the RNNE model and the winner's curse model. The decision to represent the target system by means of either of the two models involves a number of assumptions; some of them are common to both models: for instance that the rights to exploit a lease are worth the same for all bidders, or in other words that OCS auctions are common value auctions; that each bidding firm obtains an estimate from its experts; and that the estimates are unbiased, so that their mean corresponds to the common value of the track. Let us call these presuppositions common to both models 'neutral assumptions'. Other assumptions, such as the one put forward by the RNNE model that bidders are perfectly rational, are specific to either one or the other of the two models.

The role of the experiment is to discriminate between the two representations. The theoretical models provide thus the necessary information to design an experiment such that, ideally, (*a*) the neutral assumptions are satisfied, and (*b*) the experimental system can be manipulated so as to obtain a result discriminating between the two rivals, i.e. a result that can follow from only one of the two models. If so, the experiment will provide a crucial test between the two competing explanations. The process of 'demonstration' is therefore carried on in parallel, by manipulating and letting the three machines (the two theoretical models and the experimental system) run. A first process of interpretation is needed to link the theoretical models' predictions to the outcome of the experiment. But this is not the end of the story: the experimental system is not the target system, and thus one needs a further step, from the experimental auction to the real OCS auction. We shall see (sections 5.6-5.7) that this parallelism step is far from trivial.

Margaret Morrison has argued that mediating models sometimes "even supplant the physical system they were designed to represent and become the primary object of inquiry" (1998b, p. 8). This fits very well with my analogy: experimental systems too 'take on a life of their own', although their properties are investigated under the hypothesis that they can teach us something about the real economic world. But (both in the case of models and of experiments, one supposes), this applies only to the first phase of investigation. Experiments take on a life of their own, but only for a while. Because experiments in economics are separate from the target systems they intend to represent, the parallelism problem arises. Thus, good experiments must be autonomous from their target domain *only in part*, if they want to teach us something about it. In the next section I shall show how one can deal with parallelism by turning back to the OCS case.

5.6. Tightening the bridge

Demonstrating with models and experiments is an activity with a clear epistemic goal. The main question one has to face when dealing with a conjectural (theoretical or experimental) 'demonstration' of a certain phenomenon is: What is the relationship between the model or the experiment and the target system? How can one argue that the model or experiment is a satisfactory representative of the target system? Here lies the epistemic gap to be filled by parallelism arguments.[236]

It should be stressed that the problem of generalising laboratory results is not peculiar to economics. In *Laboratory Life* (1979/1986), Bruno Latour and Steve Woolgar give a detailed account of the research on the thyrotropin releasing hormone (TRH) that led to the award of the Nobel prize for medicine to Roger Guillemin in 1977. TRH is a product of the mammalian hypothalamus which allows the secretion of a hormone called thyrotropin. The hypothalamus secretes only a tiny amount of TRH, which is therefore very difficult to obtain. Guillemin in Houston (and, independently, Andrew Schally in New Orleans) solved the problem by *synthesising* a tri-peptide whose behaviour resembles very closely that of the TRH. The focus of Latour and Woolgar as microsociologists of science was on the (alleged) strategies of negotiation that led to identifying the synthesised protein with the *real* TRH in mammals' hypothalamus. The remarkable aspect of *Laboratory Life*'s case study, as noticed by Hacking (1988), is that the artificial peptide became quickly - and without much scrutiny, a fact suggesting that sociological factors may have played a role in such a decision - the benchmark for deciding what should count as TRH and what should not. Its structure became '*the*' structure of TRH. The success of Guillemin's and Schally's work was probably due to their ability to provide an 'off-the-shelf' substance potentially to be used for a number of tasks (independently of its being 'the' original TRH or not). But how can one *prove*

[236]Of course not all 'demonstrations' are put forward for epistemic purposes. When Phillips built his famous analogic model of a Keynesian macroeconomy made of pipes, water and valves, he also aimed at clarifying macroeconomic theories for didactic purposes (cf. Morgan and Boumans, 1998). His aim was pedagogic, whereas an experimental demonstration like Kagel and Levin's, in contrast, clearly has a more ambitious goal. Economic experiments are also used to convince laymen of the correctness of certain policies.

that an 'artificial' result in the laboratory is actually a faithful representative of a real-world phenomenon or entity?

Let us go back to the winner's curse experiments presented in section 5.2 above, and see how the issue was tackled in that case. Kagel and Levin claim they have produced the winner's curse phenomenon in the laboratory, and their claim is clearly intended to bear on the real-world issue at hand: *was the winner's curse the cause of average low profits in the OCS auctions?* We had two possible explanations of the data, and the experiment was designed so as to increase the plausibility of the winner's curse explanation. That is, the experiment was performed in order to confirm either the standard RNNE model, or the alternative winner's curse model (or possibly neither one nor the other) as explanations of the target phenomenon at stake.

As we have seen in the last chapter, the parallelism step must eventually be *empirical* in character. We have also seen that some experimentalists seem to be aware of this. According to Vernon Smith, for example, "which kinds of behaviour exhibit parallelism and which not can only be determined empirically by comparison studies" (1982, p. 267, but see last chapter for more quotes). In the preference reversals case, one last step was missing: having shown that certain circumstances considerably reduced the relevance of the phenomenon in the laboratory, economists did not go on to trying to answer the (crucial) question whether such conditions are instantiated in real world economies always or at least for the most part. They never, in other words, stepped out of their laboratories to collect field evidence for parallelism.

The case of the OCS auctions similarly required one step beyond the reproduction of the winner's curse in the laboratory. Real-world evidence did not play any role in the arguments presented in section 5.2 (except of course as a motivation for the experiments). Thus, the strength of the winner's curse explanation had to be somehow increased; Kagel and Levin focused on an interesting parallelism between laboratory results and a real-world phenomenon. Mead, Moseidjord and Sorensen (1983) had provided data about different profits achieved by oil companies on so-called 'wildcat' as opposed to 'drainage' leases. The former are on tracts for which no evidence about past productivity is available, whereas the latter are on tracts lying adjacent to some hydrocarbon reservoir. The developers of the adjacent tract (the 'neighbours') have higher private information on the profitability of the drainage tract, but all bidders ('non-neighbours') know that something is likely to be found.

Mead, Moseidjord and Sorensen (1983) noticed that in the Gulf of Mexico from 1954 to 1969 both neighbours and non-neighbours have had on average higher rates of returns on drainage than on wildcat leases, a fact that is not compatible with the RNNE explanation. Why? Because in RNNE, depending on whether the information available is (a) purely public, (b) purely private, or (c) both private and public, we should expect rates of return: (a) lower for all; (b-c) higher for neighbours than non-neighbours, with the latter earning less than they would in absence of insider information. If a winner's curse effect is present, in contrast, the data can be easily explained: the increase in insiders' information helps to reduce the winners' overestimation of wildcat tracts, and thus raises the returns of both neighbours and non-neighbours. Kagel and Levin (1986) show that a phenomenon of the above sort can be replicated in the laboratory, where one can control for public information at will (the strategy has been outlined in section 5.2 above).

From a methodological point of view, the logic of the procedure can be analysed as follows. Let us call the evidence in need of explanation, i.e. the fact that oil companies in the Gulf of Mexico experience on average low returns from their leases, *e*. The goal of the experiment is to discriminate between two alternative theoretical hypotheses H_1 and H_2 - the RNNE explanation and the winner's curse explanation respectively. The construction of an artificial common value auction system enables us to test (new) predictions from H_1 and H_2. Kagel and Levin, by varying initial conditions such as public/private information and the number of bidders, construct an independent test which is moreover a *quasi*-crucial experiment relative to H_1 and H_2, i.e. such that $H_2 \Rightarrow e'$ but $H_1 \Rightarrow \sim e'$. The new evidence *e*' collected in the laboratory confirms that a winner's curse phenomenon is likely to be hidden behind experimental bidding. Indeed, Kagel and Levin produced two *quasi*-crucial experiments ('*quasi*' because their conclusion depends on the decision to limit the analysis to H_1 and H_2, and on a number of auxiliary hypotheses about the experimental techniques, etc.), by varying the number of bidders and the nature of the information provided. The results of both tests were consistent with the winner's curse hypothesis, which was therefore highly confirmed.

The experimenters were aware that such evidence (*e*') could not settle the dispute about the target system. Therefore, they showed that in the real world there are cases of variation of public/private information theoretically analogous to those reproduced in the laboratory. In the OCS case, such evidence was provided by Mead, Moseidjord and Sorensen's study. The crucial argument for parallelism consists in showing that (i) in some cases the initial conditions of the real systems under study are (qualitatively) similar to those of the laboratory systems; and thus (ii) some data in the real world, $e*$, can really play the same role as e', so that $H_2 \Rightarrow e*$ and $H_1 \Rightarrow \sim e*$.

It must be noticed that in this case the data ($e*$) used to argue for parallelism constitute a sub-sample of the data (e) that originally motivated the experimental investigation. They are all, in fact, productivity data ranging over the same period. They are supposed to provide novel information nonetheless, because, although the same mechanism (either that described by the RNNE or the winner's curse mechanism) is supposed to be at work in all auctions, slightly different effects follow from different initial conditions (e.g. the kind of information available to bidders). Theoretical reasoning plays an important role in selecting the 'right' sub-set of data. This 'nesting' of e and $e*$ at any rate does not have to hold in general: the field data used for parallelism might be totally independent of the field data that motivated the experiments. When the data-set changes, other problems with generalising, inferring, inducing from one case to another necessarily arise. The inferences, once again, can only be supported by other assumptions: that the data generating process stays fixed, for example. Such assumptions may be based on empirical evidence on their own; or simply be considered reasonable in the absence of any proof to the contrary; or supported by theoretical reasoning. Background knowledge of this sort, is indispensable in the game of science.[237]

It is really $e*$ that provides confirmation for H_2 as an explanans of e. It can do so because a final process of 'interpretation' (to use Hughes' terminology) has taken place: economic theory suggests that wildcat and drainage leases provide information of a different quality (public vs. private), so that both the results 'demonstrated' from the theoretical models and those 'demonstrated' from the

[237] Moreover, the correspondence between experimental and field evidence is strictly speaking of the 'phenomena-to phenomena' kind (to take Bogen and Woodward's, 1988, terminology). To calculate profits is rather straightforward in the experiment: it is the difference between the value of the auctioned item (x_0) and the price paid for it (b_1) (this is why one does experiments in the first place!). In the field things are more complicated; Mead, Moseidjord and Sorensen (1983) use the so-called 'internal rate of return' (IRR) measure: the rate of discount which makes the present value of the stream of net revenue (i.e. gross revenue minus costs minus taxes) equal to zero. The IRR can only be estimated, because there do not exist direct data for a number of costs (exploration costs, post-sale exploration, drilling, development, production, interests, and abandonment costs, for example) which must be derived from other indicators, and the taxes attributable to each lease can only be calculated on the basis of estimated costs and revenues. Quite a number of assumptions carry the weight of the parallelism argument.

experiment can denote a phenomenon in the target system. It is crucial that such a process of interpretation, like the initial one of denotation, be 'neutral' with respect to the two theories at stake, so that the parallelism inference can be accepted by both parties. No assumption incompatible with any one of the two hypotheses can be used in order to identify *e** with *e'*.[238]

The moral is that experiments can help just at an intermediate stage of confirmation. They cannot completely fill the gap between the target phenomenon and the theoretical model. Experimentalists are aware of this:

> Our objective here is not to definitively test between competing explanations using field data. If we thought the field had this kind of potential, there would be no need to resort to laboratory experiments in the first place [...]. Rather, our objective is to show that a reasonable analysis of the available data does not falsify the hypothesis that similar economic processes are at work in both settings. If this can be done, the burden of proof rests on those who would argue that the results don't generalize to demonstrate that their arguments are correct (Kagel and Levin, 1986, p. 914).

Notice that no claims about having shifted the burden of proof are put forward before this last section of the paper. The need for an argument for parallelism is clear: experiments cannot, on their own, prove much about the real world. They can increase the plausibility of an explanation, but only up to a certain point. The reason is not only that a pattern of data can be explained by different theories, but that it may also be the result of different causal processes. A Duhem-Quine problem can be reduced in the laboratory by controlled testing, but establishing that a certain explanation is the right one in the (artificial) domain *X* does not prove that the same process lies at the origins of a similar pattern of data in the target domain *Y*. In order to establish this, one needs some further independent evidence from the target domain of application of the theory at stake. By presenting such evidence, Kagel and Levin made the first move and sent the ball into the opponents' camp. It was their turn, then, to discredit Kagel and Levin's results by challenging the parallelism argument.

5.7. Parallelism as analogy

[238] Cf. what said in section 5.5. In this sense, parallelism arguments are theory-dependent.

With the OCS example in mind, we can now try to put forward a more abstract characterisation of the relationship between models, experiments, and target systems. We can also try to define more precisely what kind of reasoning is involved in making the parallelism step from experimental systems to the real world of phenomena the theory aims at explaining. I have argued earlier that the demonstrative capacity common to models and experiments is due to their having some internal mechanism which can be 'triggered' and let run. The procedure of experimental confirmation in the non-laboratory sciences can therefore be thought of as a demonstration carried on in parallel, on three systems at the same time: a real-world system, a theoretical model, and an experimental system.

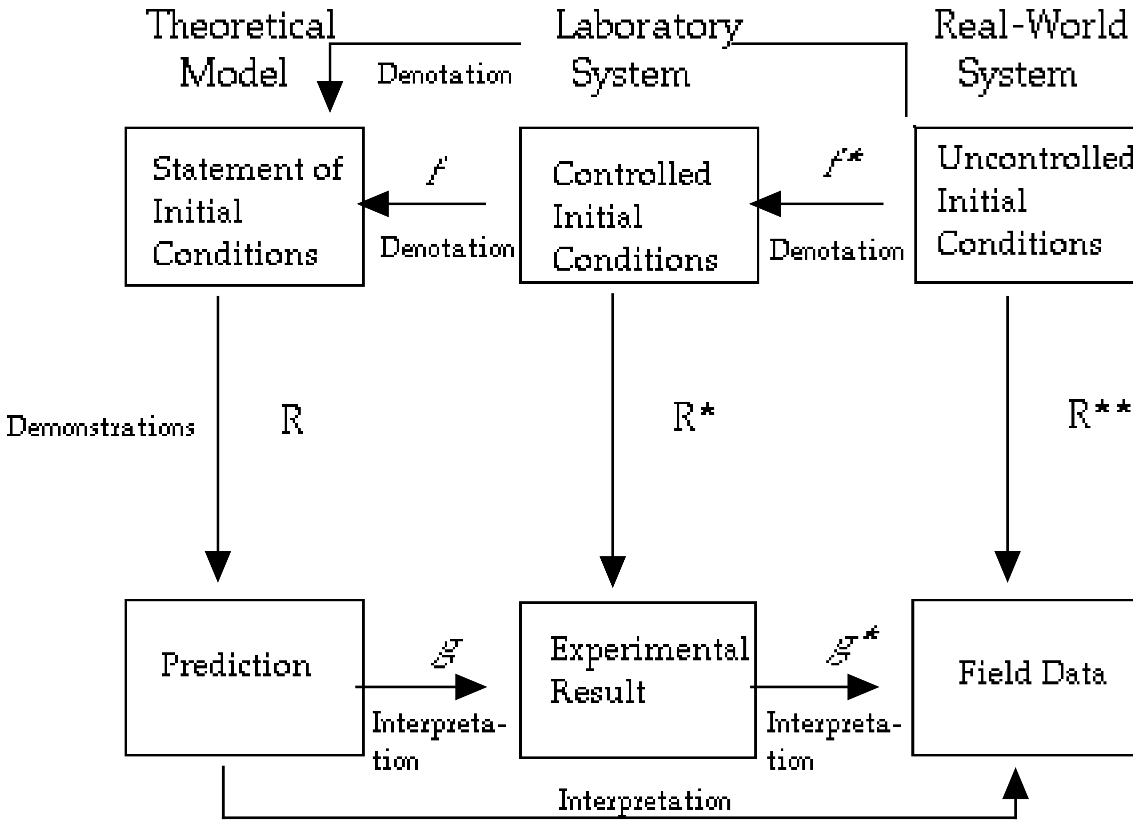

Figure 4: Three demonstrations in parallel

On the far left of figure 4, we have a (theoretical) model which, when certain initial conditions are assigned to it, leads to a prediction. On the far right, there is the real-world system, producing certain observed ‘field’ data out of naturally evolved initial conditions. In the middle, we have a laboratory system, producing an experimental result out of tightly controlled initial conditions. In order to argue for parallelism, a number of moves have to be made. The first one consists in associating the initial conditions stated by the theoretical model with both the initial conditions of the experimental system and those of the target system. The first step, as I have already suggested, is automatically fulfilled by designing the experiment so as to mirror (some of) the model’s assumptions. The second step is more problematic, as it amounts to finding some features of the real-world systems that correspond to the model’s *and* the experiment’s initial conditions. The same operation must then be carried on at the level of the outcomes of the demonstrations: the model’s predictions must be associated with the experimental outcome and with some real-world observed data.

The next (big) step consists in arguing that *since* a correspondence has been established both at the level of the initial conditions *and* at the level of the outcomes, *then* there also exists a correspondence at the level of the internal processes. Imagine two sets X and X^* such that each one element of the first set can be associated to one (and only one) element of the other by means of a function f (a bijection). Nothing guarantees that all the relations holding between the elements of the first set will also hold between the corresponding elements of the second set. In order for a relation R on the first set to translate into a relation R^* on the second set, we must impose the further requirement on the function f (taking us from the set X to the set X^*) that $x^*R^*y^*$ if and only if xRy (where x, y and x^*, y^* are elements of X and X^* respectively).

By associating some systems in such a way - claiming that the stated initial conditions of a theoretical model denote the initial conditions of an experimental system, and that both denote the initial conditions of a real-world system - we are thus drawing a function from the properties of one kind of entity to the properties of the other. The relationships exploited in parallelism arguments can be seen as functions from models to experiments to target-systems. The parallelism step is however complete only after a correspondence has been established at the level of the internal processes - the relations between a system’s initial conditions and its successive states. These relations are syntactical rules in the case of a syntactical entity, abstract relations in the case of a theoretical model, causal processes[239] in the case of an experimental and a real-world system.

To take another example: arguing from a correspondence between initial conditions and outcomes to a correspondence between mechanisms is like arguing from the observation that six points form a grid of the kind represented in figure 5 (see next page), to the claim that there exists an infinite number of other, unobservable, points forming two parallel lines. What kind of argument is this? As we have seen, Robert Sugden (fothcoming) claims that the step from a model to the real world is inductive in character. More precisely, it seems to be an analogical argument, both in the vague sense of the word 'analogy' in everyday language, and in the technical sense of the Greek word *analogìa* ('according to a ratio'). In general, an analogy is a similarity relationship between two entities or sets of entities. In the Pythagorean tradition, more precisely, an analogy was an identity of ratios. This meaning survived in the mathematical sense of analogy as proportion: $a : b = c : d$. Whereas an analogy in the everyday sense involves two entities, an analogy in the original, rigorous sense always involves at least four terms taken in couples: "As *A* is to *B*, so *C* is to *D*", according to Aristotle (*Topics*, i, 17). In the case we are concerned with, the parallelism argument amounts to an analogy of the following sort: given (*a*) some controlled initial conditions in the laboratory and (*b*) some observed experimental result on the one hand; and given (*c*) some observed properties of the target system and (*d*) some observed field data on the other, *then* (by analogy) *c* stands in the same (causal) relation to *d* as *a* stands to *b*.

[239]Of course, I do not (and cannot) go into the problem of causation here; the above claim is supposed to be neutral regarding which theory of causation one subscribes to.

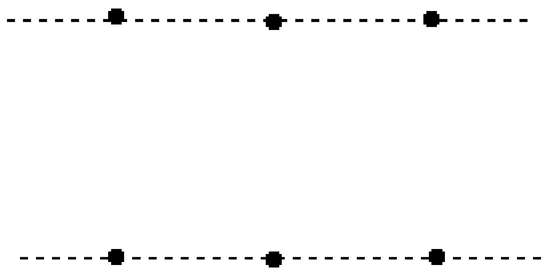

Figure 5: Parallelism as analogy

Analogies have a well-known heuristic value: by postulating an analogy between two sets of properties, we can infer the existence of a hidden property in one set by observing the existence of other properties in the other set. Analogical models sometimes work precisely this way: by observing the properties of a model we are induced to think that similar properties are to be found in the real entity modelled. To say it once again with Aristotle, "*A* is in *B* like *C* is in *D*". More rigorously, in mathematics knowledge of three terms of a postulated proportion like $1 : 3 = x : 6$ allows to obtain the value of $x = 2$. In our case, however, the parallelism analogy cannot be simply postulated. It is a hypothesis, one has to justify it, and the argument takes the form of a generalisation from a number of correspondences between the entities in the two sets to an analogy between the relations holding inside the sets themselves.

Induction is part of the process: it intervenes because, of course, the more initial conditions and outcomes are found to correspond to each other by denotation and interpretation, the more one is led to think that the systems' internal mechanisms correspond to one another. This is nothing new: it is one of the confirmation processes inductivists have taught us about since a long time ago.[240] The relationship between models, experiments and real systems is one of analogy; the analogy is established (or just weakly confirmed) by inductive reasoning.

The inference, once again, is always affected by underdetermination. Analogical reasoning may turn out to be wrong, but there is nothing special with economics, from this respect. The mechanism of propagation of poliomelytis, for example, was unknown to virologists until quite recently. In order to investigate it, Flexner and Lewis studied the process of infection in rhesus monkeys. Apparently, monkeys are easily infected via the nose, from which the virus travels to the olfactory nerves and finally to the spinal cords. Nasal sprays based on alum, zinc sulfate and picric acid seemed to be able to kill the virus, and were therefore tested in experiments on humans. But the human nervous system is less susceptible to poliomelytis than that of lower primates; on the contrary, our intestinal tract is weaker and is attacked easily by the polio virus, whereas

[240] On the problem of confirming analogies, cf. Mary Hesse's (1963) classic on models and analogies. What is said about models there can be easily translated in terms of experimental systems.

monkeys are more resistant to this kind of infection.[241] Two different causal mechanisms in this case led to the same effects, and experimental analogy proved to be misleading.

A characteristic feature of the parallelism step from experiments to target systems is that it is sometimes supported by very few established correspondences at the level of initial conditions and outcomes. In the OCS case, as we have seen, the analogy is based on just *one* correspondence - the one between private vs. public information and wildcat vs. drainage leases on the one hand, and the one between high vs. low returns of neighbour vs. non-neighbour tracts on the other. In contrast, the analogical step from the model to the experimental system can be supported by more evidence, and is thus more tightly established thanks to the manipulations and controls allowed by the laboratory. In the laboratory one can control the initial conditions so as to derive ('demonstrate') new phenomena, ideally ones that can discriminate between two alternative explanations. In the field, this is not always possible. For example, in the OCS case, field data provided a variation of public and private information but no control on the number of bidders was possible.

Are experiments needed at all, then? After all, the whole investigation started with field data and finished with field data. And, as I have argued at length, field evidence is *needed* in order to argue for parallelism. So why not go back to econometrics and recognise that that is all we have got (although it may not be very much)?[242] This would be too quick: there are good reasons to do laboratory experiments - even though these may not be *conclusive* ones in the context of a scientific controversy.

Capen, Clapp and Campbell's (1971) claim that oil companies' poor profits originated from a winner's curse phenomenon challenged the only existing theory of auctions available at the time. It also challenged a core hypothesis of mainstream economics - that economic agents are perfectly rational. According to orthodox economic theory, a winner's curse should simply not occur. Showing that it occurred in the laboratory, then, proved that it was possible. Of course the laboratory is not the real world, but in absence of a reason why the same should not occur in the target system Kagel and Levin surely proved a point - something like proving an existence theorem in mathematics. Secondly, it is true that

[241]Cf. LaFollette and Shanks (1996, pp. 126-128); for a more detailed account of this episode, see Paul (1971).

[242]I must thank in particular Julian Reiss and Peter Lipton who have raised this question.

experiments take one only half way towards the target, but they can support a point much better than most field data can. To repeat what I said in the last chapter: it is better to take one small firm step and another small shaky one, rather than one big shaky step at once.

Chapter 6

Conclusion
Simulating experiments, experimental simulations

'Isn't the same thing always true?'
'Your meaning?'
'You always have the three techniques - use, manufacture, and representation.'
'Yes.'
'And isn't the quality, beauty and fitness of any implement or creature or action judged by reference to the use for which man or nature produced it?'
(Plato, *The Republic*, Part 10, 601c-d)

7.1. Summary

Let me now summarise the main arguments of this dissertation. To begin with, I have tried to show how experiments can be used to assess normative theories. That was the central theme of chapter two: that economics is a complex discipline, concerned with matters of fact as well as normative issues such as rational decision making. To falsify an economic theory, therefore, may require an attack on both the descriptive and the normative fronts. Chapter three was devoted to provide more detail concerning the intricacies of laboratory testing. Focusing on the preference reversal experiments, I have tried to show how the interpretation of laboratory observations always requires some inference from 'data' to 'phenomena'. These inferences can be and often are indeed questioned, but they can in principle be resolved by means of ingenious testing. I then turned to the problem of parallelism and attempted a conceptual analysis of its main features in the fourth chapter. Parallelism concerns the relationship between laboratory and real world systems, and - in the light of my analysis - should be resolved on empirical grounds. In chapter five I have tried to illustrate by means of an example how parallelism can be achieved in practice. The notion of ‘mediating experiments’ has been also introduced in that chapter, and put at work on the winner’s curse example. In some circumstances, it has been suggested, parallelism can be established on firm grounds, provided enough information is available about the laboratory and the target systems in question.

This journey through economic experimentation has been motivated mainly by methodological concerns. How are experiments used? Is there any difference between the use of experiments in economics and in the natural sciences? What kind of knowledge can be gained in the laboratory? And what makes a good economic experiment? What knowledge is needed in order to apply experimental results outside the laboratory? I have tried to show that economic experiments do not differ sharply from those performed in the natural sciences. The same methods, in particular, are applied in order to detect a phenomenon of interest on the background of unreliable, ‘noisy’, or uncertain data. I have also tentatively suggested that the problem of parallelism is not specific to economics or the human sciences, but that it may hold in other contexts as well. The truth is that parallelism is often taken for granted (and sometimes for good reasons) in some disciplines, whereas, in others, experimentalists are constantly required to prove that what they are doing is relevant to understanding the ‘target phenomena’ of their discipline.

In this last chapter, finally, I would like to take a different perspective on experimental economics. We have seen what one can do with laboratory experiments, but *what is* an experiment in economics?

7.2. Experiments and simulations

Vernon Smith’s “An Experimental Study of Competitive Market Behavior”, published in 1962 in the *Journal of Political Economy*, is one of the pioneering and most influential articles in experimental economics. Throughout the paper, Smith refers to his laboratory tests as ‘experiments’, ‘experimental games’, or similar locutions. Yet, we can read in the first page that “most of these experiments have been designed to *simulate*, on a modest scale, the multilateral auction-trading process characteristic of [...] organised markets”, and Smith is keen to emphasise that “[the experiments] are intended as *simulations* of certain key features of the organised markets and of competitive markets generally" (1962, p. 111, *my italics*).

Smith's caution may appear outdated: surely economists nowadays are not afraid of speaking of *experiments*, and put their work on the same level as that of experimental physicists. But the issue is not merely terminological, and bears on an interesting conceptual distinction. Simulations, after all, are somehow supposed to be somehow less reliable than 'genuine' experiments as tools to produce scientific knowledge. They are usually taken as a second best when experimentation is unfeasible, and their results are often qualified as 'mere' simulations not to be mistaken for the 'real thing'. But economic experimental systems are not the 'real thing'. They are mediators, independent systems the study of which is supposed to teach us something about some other target systems of interest. And this sounds like the definition of a simulating device: some system (e.g., a computer-based model) that once put at work can help us understand the functioning of another system (a biological one, for example).

Of course there is little prospect for progress unless we define what experiments and what simulations are in a more precise way. Simulations have attracted the interest of philosophers of science only recently,[243] and we still lack an analysis of the distinction at stake. To my knowledge, the best solution can be found in Herbert Simon's work on the 'sciences of the artificial', but I shall come to that later. Let us begin with the opinion of someone who has taken the question of the nature of economic experiments at face value.

Harry Collins is one of the few students of experiment who have written about economic experimentation. In a short comment to a paper by Vernon Smith, Kevin McCabe and Stephen Rassenti (1991), he has argued that economic experiments are not 'genuine' experiments, of the same sort, say, as those performed by physicists. They should rather be seen as *simulations*. "Simulations", according to Collins, "have a lot in common with experiments", and certainly "*involve* a lot of experimentation and manipulation"; but "not all experimental manipulations and complex observations are experiments" (1991, pp. 227-228). Those taking place in economists' laboratories, in particular, are not.

[243]Cf. the symposium in *PSA 1990* (papers by Humphreys, Laymon and Rorlich), and more recently Hartmann (1996). Some history of the rise of simulation techniques in physics can be found in Galison (1997). As far as I know, the only historico-philosophical discussion of simulation techniques in economics is in Boumans (1998).

Collins points to some of the features common to mediators (to use the earlier terminology), such as that of being easily manipulable and carefully observable. The difference between experiments and simulations is supposed to lie elsewhere: in a genuine experiment, according to Collins, the scientist interacts with his objects of interest in a direct manner. Or perhaps we should say '*almost* direct': the physicist Robert Millikan during his famous experiments to measure the charge of the electron "did not look directly at electric charges", but rather at "the positions of the knobs that controlled the charge on the electrostatic plates", and the long journey from there to the electron "goes via oil drops, gravitational pulls, microscopes, and so on" (Collins, 1991, p. 228). Where is the difference, then?

> I think that if there is a difference it must be to do with causal chains. One can tell a plausible story about the causal relationship between charge and the position of the knobs in the case of Millikan, whereas there is no causal relationship between the world and the results of a simulation except via the *imagination* of the simulator. [...]
>
> There does not seem to be a causal connection between economies and Smith, McCabe and Rassenti's observations. If you want to tell a causal story about the relationship between the things that happen in economies and the outcome of laboratory experiments it still has to be via the experimenter's *imagination*. The experimenter has to *imagine* the relationship between economies and what happens in the lab. The economy is not to be found in the lab in the way the charge is to be found in Millikan's apparatus. The direct causal story starts with people in labs playing a game (Collins, 1991, pp. 228-229, *my emphasis*).

I have italicised Collins' repeated use of the term 'imagination', for it clearly has to do with the notion of parallelism. In the jargon used so far, Collins is pointing out that a parallelism hypothesis or assumption bears the weight of the inferences from the laboratory to the real economic world. Collins, however, uses a misleading analogy when illustrating the difference between experiments in physics and in economics. In the physics case, he focuses on the causal chain linking *laboratory* (unobservable) *entities* with *observable data*, rather than laboratory entities with *non-laboratory* ones. But the latter question can in principle be raised also in the Millikan case: what is the relationship between Millikan's laboratory electrons and real-world electrons? The relationship between the charges measured in his tightly controlled artificial environment and those of the electrons, say, in my chair? Millikan's experiment was a good one if it isolated (or we better say *reproduced*, given the complex preparation of materials required by such experiments) in the laboratory electrons with the same properties (among those of interest, like charge) as those to be found elsewhere, in the real world of unconstrained microparticles. But the same applies to economics: the fact that the step from the laboratory to the outside world is usually taken for granted by physicists does not mean that such a step is not taken. Parallelism may be invisible, but is always there.

The distinction between simulations and experiments does not have to do with the observable vs. unobservable distinction, nor with the 'internal' vs. 'external' validity problem. Both hold (at different degrees, perhaps) in all sciences. Collins has the right intuition when pointing to causal chains as the key to this puzzle, but frames the question in the wrong fashion. Let us try a different approach, then.

Herbert Simon once noticed that simulations rely on a process of abstraction from the fundamental principles governing the behaviour of the simulating and the target systems (1969, pp. 15-18). One starts from the hypothesis that the same 'organisational properties' arise at a certain non-fundamental level from different substrata. If the hypothesis holds, it is legitimate to abstract from the latter and simulate the behaviour of a system *A* by observing the behaviour of another system *B* which happens to (or which is purposely built so as to) display these non-fundamental properties.

Working on this idea, we can now propose a criterion to demarcate genuine experiments from 'mere' simulations. The difference lies in the kind of similarity relationship existing between, on the one hand, an *experimental* and its *target* system, and, on the other, a *simulating* and its target system. In the former case, the similarity is supposed to hold at a 'deep' and 'material' level, whereas in the latter case the similarity is admittedly only 'abstract' and 'formal'.

A common feature of all mediating entities is their greater manageability, simplicity, and controllability than that of the target systems they are supposed to help understand. A mediating entity may be used to study the functioning of some object that is too complex or can hardly be manipulated, but only if it resembles the target system in some relevant respect. At a very general level, two systems may be similar in the sense that some abstract relationship holds in both of them. For example, take the equation

$$m\frac{d^2x}{dt^2}+\beta\frac{dx}{dt}+kx=F(t).$$

It describes the motion of an ideal mechanical system with one degree of freedom (like for instance a sphere attached to a spring): $F(t)$ is the acting force, x is the motion of the mass m, k the elastic constant of the spring, and β a friction coefficient. A formally identical equation is used to describe an electric charge moving in a section of electric circuit made of an autoinduction in series with a resistor and a condenser. In this case we have in fact

$$L\frac{d^2Q}{dt^2}+R\frac{Qx}{dt}+\frac{1}{C}Q=V(t),$$

where L is the autoinduction, R the resistance, Q the charge of the condenser at time t, and $V(t)$ the voltage in the circuit. The abstract electric circuit can be taken as an *analogical* model of the abstract mechanical system constituted by a little sphere attached to a spring - because both are legitimate interpretations of the above equations (or rather: *equation*, singular). Tension is the analogical counterpart of force, electric current is the counterpart of velocity, charge is the counterpart of displacement, autoinduction is the counterpart of mass, and so on. We are here dealing with two theoretical models which are analogous at the formal level. The interpretation of the terms in the equations is different, depending on which model is chosen to interpret the mathematical relationship.

As already pointed out in chapter five, theoretical models usually do not exist physically. They are mostly abstract entities 'picked up' by implicit definition by some set of sentences (the syntax of the theory). Theoretical models can however sometimes be used to construct concrete, material ones. Engineers, for example, often build scale models of cars, ships, and airplanes for testing purposes. Sometimes economists do it too: the famous 'Phillips machine' is a concrete model of a Keynesian economy made of pipes, valves, and coloured liquid, built for pedagogical purposes in order to show in a pictorial way the flow of income in a macroeconomy.[244] This is however a peculiar case, and I shall here illustrate my point by relying on a physics example.

A material model of the propagation of light, according to the wave theory, can be built with the aid of water waves in a ripple tank. At a rather general level

---

[244]See Morgan and Boumans (1998) for some history and methodology of the Phillips machine.

of analysis, in fact, any kind of wave can be modeled as a perturbation in a medium determined by two forces: the external force producing the perturbation, and the reacting force trying to restore the medium at rest. General relationships such as Hooke's law or D'Alembert's equation may hold for *all* kind of waves. More fundamental relationships, such as Maxwell's equations, describe the properties of the electric and the magnetic field only. The values given by Maxwell's equations can be used in D'Alembert's wave equation in order to obtain, for instance, the velocity of propagation of an electromagnetic wave, because electricity behaves *like* a wave, although the fundamental principles at work are different from those at work in case of, e.g., water waves. The terms appearing in the equation describing the target and the model-systems are to be interpreted differently in the two cases: the forces at work are different in nature, and so are the two media in which waves propagate. The similarity between the theoretical model of light waves and the ripple-tank model holds at a very abstract level only. The two systems are made of different 'stuff': water waves are not light waves.

Because of this formal similarity, though, the behaviour of light waves can be simulated in a ripple tank. Both light waves and water waves seem to obey the same non-structural law, despite their being made of different 'stuff'. This is due to different reasons in each case: different underlying processes produce similar behaviour at a more abstract level of analysis.[245] Similarly, human behaviour can to some extent be simulated with computerised models, but the former arise from 'machines' made of flesh, blood, neurons, etc. rather than silicion chips. The Aristotelian distinction between 'material' and 'formal' causes seems to provide an appropriate framework for the present discussion:[246] according to Aristotle one can 'explain' a given statue by means of its form, but also in terms of its material constituents, e.g. its being made of marble. This latter explanation is different from the 'formal' one in that marble itself may in principle be explained in terms of its own formal causes. Yet, for the task at hand (explaining the statue) one does not have to go as deep as that in the formal structure of the object under study. The material can be abstracted from for pragmatic reasons. The same with simulations: they are particularly valuable when one cannot manipulate nor

[245]Of course, if one believes in the reductionist story according to which everything physical is made of the same fundamental sub-atomic particles, then both light and water waves are 'made of the same stuff'. But the reductionistic story is controversial (photons for example seem to have different properties from other particles), and at any rate the fact that everything is made of the same stuff does not play any relevant role in explaining why both systems display certain non-fundamental relations.

[246]For an extended discussion of the Aristotelian point of view, cf. McMullin (1977).

perform controlled experiments on *A*, and thus looks for an analogous system *B* similar only up to a certain degree, in order to investigate or simply illustrate certain specific properties of *A*.

7.3. Experimental simulations

Let us now return to economics. The crucial presumption behind mediating experiments is that some relevant components of the laboratory system are made of the same 'stuff' as those of the target system. The experiment should feature the same causal processes that are at work in the real world, rather than display some abstract relationships by means of a different device. If the causal principles at work are different in the laboratory and in the target system, then the experiment is simply a bad one. It is a failed experiment, not a good simulation.

This distinction helps to account for the advantages of laboratory experimentation, its use, and its being complementary to simulating techniques. Experiments are particularly useful when one has an imperfect understanding of the basic causal mechanisms of the system under study. They can be used in these contexts because the laboratory 'stuff' is assumed to be the same as the non-laboratory 'stuff'. For example, one can do experiments on market behaviour even without a proper understanding of the mechanisms of individual choice and belief formation. Experimental subjects may trade at a certain equilibrium price because they are acting in a fully rational way, or perhaps because they are following some rule of thumb, or even by sheer imitation. Whatever the real causal process, we can use laboratory tests to study the functioning of specific real world economies as long as we are confident that the same (unknown) basic principles of behaviour apply in both cases.

An experiment can give us more confidence in a model, only if the model makes some contestable assumption about some component of the target system, and if the experiment includes the real component (for example, real human behaviour, las in Kagel and Levin's experiments). An experiment that merely reproduces the assumptions of the model, for example paying subjects to act according to the behavioural theory of the model, does not test anything at all – except perhaps the incentive system.

Considerations of this kind have often been proposed by experimentalists to defend their methodology. Vernon Smith, for example, puts forward a double argument in support of his faith in the relevance of laboratory experimentation.

"The laboratory becomes a place where real people earn real money for making real decisions about abstract claims that are just as 'real' as a share of General Motors". For this reason, "Laboratory experience suggests that all the characteristics of 'real world' behavior that we consider to be of primitive importance [...] arise naturally, indeed inevitably, in experimental settings" (1976, pp. 100-101). This reasoning supports experimentalists' confidence in their results. To them, the 'real' character of experimental markets helps to bridge the gap between a theory and its intended target. "Laboratory microeconomies are real live economic systems, which are certainly richer, behaviorally, than the systems parametrized in our theories" (Smith, 1982, p. 923). Experimental economies are indeed supposed to work according to the same principles as the target systems in the intended domain of economic theory, because they are made of the same 'material'.

Simulations and experiments are therefore the appropriate research tools in different contexts. Roughly, simulations can be used in two different ways: (1) either to bootstrap from the fact that a given effect (which we have first observed in system *A*) has been produced by means of simulation *B*, to the fact that the relations governing the behaviour of *B* also govern the behaviour of *A*; (2) or to argue that a certain effect observed by simulating with *B* will also be observed in the case of *A* because the two are governed by similar relations.[247] Both procedures are knowledge-producing ones. The point to be stressed, however, is that in both cases the governing relationships have to be fully specified for the simulations to be carried on. Simulations of this kind are 'transparent boxes', to which the old *dictum* applies: 'the results of a simulation are only as good as the assumptions that you feed into it'.

Geologists working in stratigraphy, for instance, study the structure of rock strata below the earth's surface. They also investigate the process of formation of strata, but usually have to face very serious obstacles, such as the impossibility of doing controlled experiments (processes of sedimentation last for millenia and of course the target systems are too large to be manageable), the difficulty to gather data even about the present geography of the strata, the strong theory dependence of the interpreted data, and the complex interdependencies within geological systems. In order to solve at least some of these problems, stratigraphists have devised simulation techniques such as STRATAGEM, a

[247]One may be unable to experiment with *A*, or the equations describing *A* may be so complicated that they can be solved only by means of some 'brute-force' solution in *B*, etc.

computer-based modeling package used also by large companies such as Shell Oil.[248] These simulation techniques work on the basis of a number of structural equations taken from the theory of 'sequence stratigraphy'. The equations model the system's outcome (the actual sedimentation) as a function of a number of variables such as the hydrodynamics of sediment deposition, the subsidence patterns, the global sea level, the amount of sediment supplied to the basin, etc. The point is that the outcome of the simulation is dependent on the approximate validity of the theory endorsed, and also on the correct specification of the initial conditions and of the values assigned to the free parameters in the equations (and these are all problematic assumptions to make in the specific case).

Contrary to Collins, it is not necessarily true that "simulations do not have the epistemological priority that we accord to experiments" (1991, p. 229). Rather, the knowledge needed to run a good simulation *is not the same* as the one needed to run a good experiment. As noticed by Hughes, "computer experimentation [i.e. simulation] is in a crucial respect on a par with all other kinds of theoretical speculation" (1999, p. 63), in that it requires empirical confirmations. When reproducing a real-world system in the laboratory, similarly, the relationships which describe the behaviour of both systems are not well-known in advance. But one does not have to specify the full set of structural equations governing the target system. The trick is to *make sure* that the target and the experimental system are similar in most relevant respects, so as to generalise the observed results from the laboratory to the outside world. Experimenters make sure that this is the case by using materials that resemble as closely as possible those of which the parts of the target system are made. They also make sure that the different components of the mimetic device are put together just like those of the target, and that nothing else is interfering. Of course, quite a lot of knowledge is required in order to do so, but no fundamental theory of how the target system works is needed. The laboratory system can be partly used as a 'black box' device.

As we have seen in chapter five, Kagel and Levin claimed they had shifted the burden of proof by showing that a winner's curse phenomenon can be observed in a laboratory 'common value auction' which is analytically analogous to (that is, it shares all the characteristics that auction theory deems relevant, with) the real OCS auctions that had originally attracted economists' attention. The burden of

---

[248] I have learned about simulation techniques in geology from Francis Longworth - see his unpublished paper on the methodology of STRATAGEM.

proof was actually shifted because *there was no reason to believe that the experimental system was different from the target system of interest*. The same processes, the same causal principles were supposed to be at work in both cases - until proof to the contrary. Experimental systems thus are close to real world economies, because (when correctly designed) they are made of the same 'stuff' as real world economies. No process of abstraction from the material forces at work is needed in order to draw the analogy from the laboratory to the outside world. One may abstract from 'negligible' causal factors, but not from the basic processes at work. The similarity is not merely *formal*, but holds at the *material* level as well.

The difference between experiments and simulations can be best appreciated if one turns to 'hybrid' experiments. I call them 'hybrids' because they share characteristics of *both* 'pure experiments' and 'pure simulations' at the same time. Game theorists, for example, often impose for analytical reasons very tight restrictions on their models, by assuming for instance absolutely risk averse players. When experimenters try to test the predictions of such models, they have to make sure that such conditions are satisfied. In the particular case of risk aversion, they have two possible strategies: they can pre-test subjects and use for the purpose of the experiment only those who happen to display risk averse behaviour. Or they can use an ingenious experimental device invented by Roth and Malouf (1979): in order to 'produce' risk-averse behaviour, the experimenters use a 'reward medium' consisting of entry-tickets for lotteries. The lotteries can then be set so that the subjects are indeed maximising some 'risk-averse' (i.e. concave from above) 'medium function' rather than their own, *real* utility function. The index of preferences is in this case *simulated* by means of the medium function.

Or, to take a more familiar example, let us go back to Kagel and Levin's experiments on the winner's curse phenomenon: Kagel and Levin (1986) did not experiment on subjects who were uncertain about the value of the items to be auctioned in exactly the same way as an oil company manager is uncertain about the value of an auctioned tract. They rather provided each subject with a 'private signal' of the real value of the item, drawn from a lottery. They thus *simulated* uncertainty by means of a random draw, because auction theory models agents' uncertainty as a probability distribution of this sort. The uncertainty arose from the interaction of experimental subjects with a random lottery device, rather than with an oil tract of unknown value. In order to do so, one needs to be very confident that such a way of modeling this particular aspect of the target system under study is legitimate. One has to be reasonably sure, in other words, that *that* part of the theory is right.[249]

That feature of the experiments was, so to speak, 'hyper-realistic': in the target economy, beliefs are formed by subjects facing an information set which is definitely less neat and ready-to-use than the one provided by the experimenters. Kagel and Levin's subjects also received very quick feedback about the real value of the auctioned item, whereas in the real OCS auctions the real value of an oil tract is revealed only after many years and the process of belief revision may work differently. These experiments, then, shared some features of simulations, suggesting that there really is a continuum of cases rather than a sharp distinction. They are hybrid objects - 'simulating experiments', or if you prefer 'experimental simulations'.[250]

7.4. Mediating entities, mimetic devices

If we take the above reasoning seriously, we must conclude that there exists a continuum of *mediating entities* - with different but somehow related properties, used in different but not entirely unrelated ways. They are, to borrow Galison and Assmus' (1989) terminology, '*mimetic* devices' used to reproduce the behaviour of some other system. On one extreme of this ideal taxonomy, there are abstract theoretical models; in the middle, we find a sequence of concrete models, simulating devices, and experimental systems. At the opposite end, the target systems that lucky scientists have a chance to manipulate, dissect and reassemble - in other words, to experiment upon without any mediation.

[249]One can object (as Robert Sugden has done in private correspondence) that laboratory and real world economies have to be seen as tokens of a same kind ('the economy'). Thus, to ask whether uncertainty has been produced correctly in the laboratory would be like asking whether in an experiment to investigate the properties of steam, the boiler has been heated by burning coil rather than by a Bunsen burner: the heat is heat however it is produced, and obeys to the laws of thermodynamics. As I have argued in chapter four, however, this kind of reasoning does not seem to be warranted in economics, where universal covering laws are hard to find. If economic relationships hold on the 'right' background circumstances only, then experimental inferences have to be local in character. Of course there may be mechanisms general enough to support a wide range of inferences, but that is an empirical matter.

[250]Hartmann (1996) uses the second label.

Bibliography

Achinstein, P. and O. Hannaway (eds. 1985) *Observation, Experiment and Hypothesis in Modern Physical Science,* Cambridge, Mass., MIT Press.

Ackermann, R. (1985) *Data, Instruments, and Theory*, Princeton, Princeton University Press.

Ackermann, R. (1989) “The New Experimentalism”, *British Journal for the Philosophy of Science*, 40, pp. 185-190.

Allais, M. (1953/1979) "Fondement d'une théorie positive des choix comportant un risque et critique des postulats et axiomes de l'ecole americaine", in C.N.R.S. (1953); Engl. transl. "The Foundations of a Positive Theory of Choice Involving Risk and a Criticism of the Postulate and Axioms of the American School", in M. Allais and O. Hagen (eds.) *Expected Utility Hypothesis and the Allais Paradox*, Dordrecht, Reidel, 1979.

Allais, M. (1979) "The So-called Allais Paradox and Rational Decisions Under Uncertainty", in M. Allais and O. Hagen (eds. 1979) *Expected Utility Hypothesis and the Allais Paradox*, Dordrecht, Reidel.

Allais, M. (1988a) “The General Theory of Random Choices in Relation to the Invariant Cardinal Utility Function and the Specific Probability Function. The (U, θ) Model”, in B.M. Munier (ed.) *Risk*, *Decision and Uncertainty*, Dordrecht, Reidel.

Allais, M. (1988b) “A Neo-Bernoullian Theory: The Machina Theory. A Critical Analysis”, in B.M. Munier (ed.) *Risk, Decision and Uncertainty*, Dordrecht, Reidel.

Allais, M. (1992) “Contributions scientifiques à la théorie de l’utilité et du risque: théorie, experience, et applications”, paper presented at the 6th ‘Foundations of Utility and Risk’ conference, Paris.

Amihud, Y. (1979) "A Reply to Allais", in M. Allais and O. Hagen (eds. 1979) *Expected Utility Hypothesis and the Allais Paradox*, Dordrecht, Reidel.

Anand, P. (1993) *Foundations of Rational Choice Under Risk,* Oxford, Oxford University Press.

Arrow, K.J. (1951) "Alternative Approaches to the Theory of Choice in Risk-Taking Situations", *Econometrica*, 19, pp. 404-437.

Arrow, K.J. (1971) *Essays in the Theory of Risk Bearing*, Chicago, Markham.

Bachelard, G. (1933) *Les intuitions atomistiques*, Paris, Boivin.

Bachelard, G. (1953) *Le materialisme rationnel*, Paris, Presses Universitaires de France.

Backhouse, R.E. (1994) "The Lakatosian Legacy in Economic Methodology", in R.E. Backhouse (ed.) *New Directions in Economic Methodology*, London, Routledge.

Backhouse, R.E. (1997) *Truth and Progress in Economic Knowledge*, Cheltenham, Edward Elgar.

Barnes, B. (1974) *Scientific Knowledge and Sociological Theory*, London, Routledge.

Baumol, W.J. (1951) "The Neumann-Morgenstern Utility Index: An Ordinalist View", *Journal of Political Economy*, 59, pp. 61-66.

Baumol, W.J. (1958) "The Cardinal Utility which is Ordinal" *Economic Journal*, 68, pp. 665- 672.

Becker, G.M., M.H. DeGroot, and J. Marschak (1964) "Measuring Utility by a Single-Response Sequential Method", *Behavioral Science*, 9, pp. 226-232.

Berg, J.E., J.W. Dickhaut, and J.R. O'Brien (1985) "Preference Reversal and Arbitrage", *Research in Experimental Economics*, 3, pp. 31-71.

Bernard, C. (1865) *Introduction à l'étude de la Medicine Experimentale*, Paris, Flammarion; Engl. transl. *Introduction to the Study of Experimental Medicine*, New York, Henri Schumann, 1957.

Bernoulli, D. (1738/1954) "Specimen theoriae novae de mensura sortis", *Commentarii academiae scientiarum imperialis Petropolitanae*, 5, pp. 175-192; Engl. transl. (1954) "Exposition of a New Theory on the Measurement of Risk", *Econometrica*, 22, pp. 23-36.

Bhaskar, R. (1975) *A Realist Theory of Science*, Brighton, Harvester.

Binmore, K. (1999) "Why Experiment in Economics?", *Economic Journal*, 109, pp. F16-F24.

Blaug, M. (1980/1992) *The Methodology of Economics*, Cambridge, Cambridge University Press.

Bloor, D. (1976) *Knowledge and Social Imagery*, London, Routledge.

Bogen, J. and J. Woodward (1988) “Saving the Phenomena”, *Philosophical Review*, 97, pp. 303-352.

Boumans, M. (1998) “Lucas and Artificial Worlds”, in J.B. Davis (ed.) *New Economics and Its History*, HOPE supplement to Vol. 29, Durham, Duke University Press.

Boyssou, D. and J.C. Vansnick (1990) “‘Utilité cardinale’ dans le certain et choix dans le risque”, *Revue économique*, 41, pp.979-1000.

Buchwald, J.Z. (1993) “Design for Experimenting”, in P. Horwich (ed.) *World Changes*, Cambridge Mass., MIT Press.

Buchwald, J.Z. (ed. 1995) *Scientific Practice*, Chicago, University of Chicago Press.

C.N.R.S. (1953) *Économétrie (Paris 12-17 mai 1952)*, Colloques Internationaux du Centre National de la Recherche Scientifique.

Cambell, D. and J. Stanley (1966) *Experimental and Quasi-Experimental Designs for Research*, Chicago, Rand McNally.

Camerer, C. (1989) “An Experimental Test of Several Generalised Utility Theories”, *Journal of Risk and Uncertainty*, 2, pp. 61-104.

Camerer, C. (1995) “Individual Decision Making”, in J.H. Kagel and A.E. Roth (eds.) *The Handbook of Experimental Economics*, Princeton, Princeton University Press.

Camerer, C. (1997) “Progress in Behavioral Game Theory”, *Journal of Economic Perspectives*, 11, pp. 167-188.

Capen, E.C., R.V. Clapp, and W.M. Campbell (1971) “Competitive Bidding in High-Risk Situations”, *Journal of Petroleum Technology*, 23, pp. 641-653.

Carnap, R. (1934) *Logische Syntax der Sprache*, Vienna, Springer; Engl. transl. *The Logical Syntax of Language*, London, Kegan Paul, Trench, Trubner & Co., 1937.

Carnap, R. (1950) *Logical Foundations of Probability*, Chicago, University of Chicago Press.

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford, Clarendon Press.

Cartwright, N. (1989) *Nature’s Capacities and Their Measurement*, Oxford, Oxford University Press.

Cartwright, N. (1991) "Replicability, Reproducibility, and Robustness: Comments on Harry Collins", *History of Political Economy*, 23, pp. 143-155.

Cartwright, N. (1993) "How We Relate Theory to Observation", in P. Horwich (ed.) *World Changes*, Cambridge Mass., MIT Press.

Cartwright, N. (1994) "Fundamentalism vs. the Patchwork of Laws", *Aristotelian Society Proceedings*, 94, pp. 279-292.

Cartwright, N. (1995) "*Ceteris Paribus* Laws and Socio-Economic Machines", *The Monist*, 78, pp. 276-294.

Cartwright, N., J. Cat, L. Fleck and T. Uebel (1996) *Otto Neurath: Philosophy between Science and Politics*, Cambridge, Cambridge University Press.

Chamberlin, E.H. (1948) "An Experimental Imperfect Market", *Journal of Political Economy*, 56, pp. 95-108.

Chew, S.H. and K. MacCrimmon (1979) "Alpha-Nu Choice Theory: A Generalization of Expected Utility Theory", University of British Columbia, Faculty of Commerce and Business Administration, Working Paper no. 686.

Chu, Y.P. and R.L. Chu (1990) "The Subsidence of Preference Reversals in Simplified and Marketlike Experimental Settings: A Note", *American Economic Review*, 80, pp. 902-911.

Coffa, J.A. (1973) *The Foundations of Inductive Explanation*, doctoral dissertation, University of Pittsburgh; University Microfilms, Ann Arbor, Michigan.

Collins, H.M. (1975) "The Seven Sexes: A Study in the Sociology of a Phenomenon, or the Replication of Experiments in Physics", *Sociology*, 9, pp. 205-224.

Collins, H.M. (1984) "*When* Do Scientists Prefer to Vary Their Experiments?", *Studies in History and Philosophy of Science*, 15, pp. 169-174.

Collins, H.M. (1985) *Changing Order,* Beverly Hills, Sage.

Collins, H.M. (1991a) "Comment" [on Smith, McCabe and Rassenti (1991)], in N. de Marchi and M. Blaug (eds.) *Appraising Economic Theories*, Aldershot, Elgar.

Collins, H.M. (1991b) "History and Sociology of Science and History and Methodology of Economics", in de Marchi, N. and M. Blaug (eds.) *Appraising Economic Theories*, Aldershot, Elgar.

Collins, H.M. (1994) “A Strong Confirmation of the Experimenter’s Regress”, *Studies in History and Philosophy of Science*, 25, pp. 493-503.

Cox, J.C. and D.M. Grether (1996) “The Preference Reversal Phenomenon: Response Mode, Markets and Incentives”, *Economic Theory*, 7, pp. 381-405.

Cox, J.C. and S. Epstein (1989) “Preference Reversals Without the Independence Axiom”, *American Economic Review*, 79, pp. 408-426.

Cross, J. (1980) “Some Comments on the Papers by Kagel and Battalio and by Smith”, in J. Kmenta and J. Ramsey (eds.) *Evaluation of Econometric Models*, New York, New York University Press.

Cunningham, A. and P. Williams (eds. 1992) *The Laboratory Revolution in Medicine*, Cambridge, Cambridge University Press.

Davis, D.D. and C.H. Holt (1993) *Experimental Economics,* Princeton, Princeton University Press.

De Vroey, M. (1998) “Is the Tâtonnement Hypothesis a Good Caricature of Market Forces?”, *Journal of Economic Methodology*, 5, pp. 201-222.

de Finetti, B. (1952) “Role de la théorie des jeux dans l’économie et role des probabilités personnelles dans la théorie des jeux”, in C.N.R.S. (1953) *Économétrie* (Paris 12-17 mai 1952), Colloques Internationaux du Centre National de la Recherche Scientifique.

Duhem, P. (1906) *La théorie physique. Son objet et sa structure*, Paris, Chevalier et Rivière; Engl. transl. *The Aim and Structure of Physical Theory*, Princeton, Princeton University Press, 1954.

Dupré, J. (1993) *The Disorder of Things*, Cambridge Mass., Harvard University Press.

Edwards, W. (1953) “Experiments on Economic Decision-Making in Gambling Situations”, *Econometrica*, 21, pp. 349-350.

Egidi, M. (1995) “Routines, Hierarchies of Problems, Procedural Behaviour: Some Evidence from Experiments”, in K.J. Arrow, E. Colombatto, M. Perlman and C. Schmidt (eds.) *The Rational Foundations of Economic Behaviour*, London, Macmillan.

Ellsberg, D. (1954) “Classic and Current Notions of ‘Measurable Utility’”, *Economic Journal*, 64, pp. 528-556.

Epstein, P. (1999) "Wesley Mitchell's Grand Design and Its Critics: The Theory and Measurement of Business Cycles", *Journal of Economic Issues*, forthcoming.

Evans, R. (1997) "Soothsaying or Science? Falsification, Uncertainty and Social Change in Macroeconomic Modelling", *Social Studies of Science*, 27, pp. 395-438.

Feigl, H. (1970) "The 'Orthodox' View of Theories: Remarks of Defence as well as Critique", in M. Radner and S. Winokour (eds.) *Minnesota Studies in the Philosophy of Science*, Vol. 4, Minneapolis, University of Minnesota Press.

Feyerabend, P.K. (1975/1993) *Against Method*, London, Verso.

Fishburn, P. and P. Wakker (1995) "The Invention of the Independence Condition for Preferences", *Management Science*, 41, pp. 1130-1144.

Fishburn, P.C. (1989) "Retrospective on Utility Theory of von Neumann and Morgenstern", *Journal of Risk and Uncertainty*, 2, pp. 127-138.

Fishburn, P.C. (1991) "Nontransitive Preferences in Decision Theory", *Journal of Risk and Uncertainty*, 4, pp. 113-134.

Fleck, L. (1935/1979) *The Genesis and Development of a Scientific Fact*, Chicago, University of Chicago Press.

Fodor, J.A. (1974) "Special Sciences (Or: The Disunity of Science as a Working Hypothesis)", *Synthese*, 28, pp. 97-115.

Fodor, J.A. (1987) *Psychosemantics*, Cambridge, Mass., MIT Press.

Fodor, J.A. (1989) "Making Mind Matter More", *Philosophical Topics*, 17, pp. 59-79.

Fodor, J.A. (1991) "You Can Fool Some of the People All of the Time, Everything Else Being Equal; Hedged Laws and Psychological Explanations", *Mind*, 100, pp. 19-34.

Franklin, A. (1986) *The Neglect of Experiment*, Cambridge, Cambridge University Press.

Franklin, A. (1990) *Experiment, Right or Wrong*, Cambridge, Cambridge University Press.

Franklin, A. (1994) "How to Avoid the Experimenter's Regress", *Studies in the History and Philosophy of Science*, 15, pp. 51-62.

Franklin, A. (1998) "Experiment in Physics", in *The Stanford Encyclopedia of Philosophy*, available on line at http://plato.stanford.edu/entries/physics-experiment

Franklin, A. and C. Howson (1984) "Why Do Scientists Prefer to Vary Their Experiments?", *Studies in History and Philosophy of Science*, 15, pp. 51-62.

Franklin, A. and C. Howson (1988) "It Probably Is A Valid Experimental Result", *Studies in History and Philosophy of Science*, 19, pp. 419-427.

Friedman, D. and S. Sunder (1994) *Experimental Methods: A Primer for Economists*, Cambridge, Cambridge University Press.

Friedman, M. (1974) "Explanation and Scientific Understanding", *Journal of Philosophy*, 71, pp. 5-19.

Friedman, M. and L.J. Savage (1948) "The Utility Analysis of Choices Involving Risk", *Journal of Political Economy*, 56, pp. 279-304.

Friedman, M. and L.J. Savage (1952) "The Expected-Utility Hypothesis and the Measurability of Utility", *Journal of Political Economy*, 60, pp. 463-474.

Galison, P. (1987) *How Experiments End*, Chicago, University of Chicago Press.

Galison, P. (1995) "Context and Constraints", in Buchwald (ed.) *Scientific Practice*, Chicago, Chicago University Press.

Galison, P. (1997) *Image and Logic*, Chicago, University of Chicago Press.

Galison, P. and A. Assmus (1989) "Artificial Clouds, Real Particles" in D. Gooding, T. Pinch and S. Schaffer (eds.) *The Uses of Experiment*, Cambridge, Cambridge University Press; partially reprinted as ch. 2 of P. Galison, *Image and Logic*, Chicago, University of Chicago Press, 1997.

Galison, P. and D.J. Stump (eds. 1996) *The Disunity of Science*, Stanford, Stanford University Press.

Giere, R.N. (1979/1997) *Understanding Scientific Reasoning*, New York, Harcourt Brace, 4th ed.

Giere, R.N. (1988) *Explaining Science*, Chicago, University of Chicago Press.

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

Gooding, D. (1990) *Experiment and the Making of Meaning*, Dordrecht, Kluwer.

Gooding, D., T.J. Pinch and S. Schaffer, (eds. 1989) *The Uses of Experiment: Studies in the Natural Sciences,* Cambridge, Cambridge University Press.

Grether, D. and C. Plott (1979) "Economic Theory of Choice and the Preference Reversal Phenomenon", *American Economic Review*, 69, pp. 623-638.

Guala, F. (unpublished) "Constructing Institutions in the Laboratory: The FCC Case", CPNSS, London School of Economics.

Hacking, I. (1983) *Representing and Intervening*, Cambridge, Cambridge University Press.

Hacking, I. (1988) "The Participant Irrealist at Large in the Laboratory", *British Journal for the Philosophy of Science*, 39, pp. 277-294.

Hacking, I. (1989a) "Philosophers of Experiment", *PSA 1988*, Volume 2, East Lansing, Philosophy of Science Association, pp. 147-156.

Hacking, I. (1989b) "Extragalactic Reality: The Case of Gravitational Lensing", *Philosophy of Science*, 56, pp. 555-581.

Hacking, I. (1992) "The Self-Vindication of the Laboratory Sciences", in A. Pickering (ed.) *Science as Practice and Culture*, Chicago, University of Chicago Press.

Hacking, I. (1996) "The Disunities of the Sciences", in Galison and D.J. Stump (eds.) *The Disunity of Science,* Stanford, Stanford University Press.

Hacking, I. (1999) *The Social Construction of What?*, Cambridge Mass., Harvard University Press..

Hands, D.W. (1990) "Second Thoughts on 'Second Thoughts': Reconsidering the Lakatosian Progress of the General Theory", *Review of Political Economy*, 2, pp. 69-81.

Hanson, N.R. (1958) *Patterns of Discovery,* Cambridge, Cambridge University Press.

Hansson, B. (1988) "Risk Aversion as a Problem of Conjoint Measurement", in P. Gärdenfors and N.E. Sahlin (eds.) *Decision, Probability, and Utility: Selected Readings*, Cambridge, Cambridge University Press.

Hargreaves Heap, S. and Y. Varoufakis (1995) “Experimenting with Neoclassical Economics: A Critical Review of Experimental Economics”, in I.H. Rima (ed.) *Measurement, Quantification and Economic Analysis*, London, Routledge.

Harrison, G. (1989) “Theory and Misbehavior of First-Price Auctions”, *American Economic Review*, 79, pp. 749-762.

Harsanyi, J.C. (1977) “Morality and the Theory of Rational Behavior”, in A. Sen and B. Williams (eds. 1982) *Utilitarianism and Beyond*, Cambridge, Cambridge University Press.

Hartmann, S. (1996) “The World as a Process: Simulations in the Natural and Social Sciences”, in R. Hegselmann, U. Mueller and K.G. Troitzch (ed.) *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*, Dordrecht, Kluwer.

Hausman, D.M. (1989) “*Ceteris Paribus* Clauses and Causality in Economics”, *PSA 1988*, Vol. 2, East Lansing, Philosophy of Science Association.

Hausman, D.M. (1991) “On Dogmatism in Economics: The Case of Preference Reversals”, *Journal of Socio-Economics*, 20, pp. 205-255.

Hausman, D.M. (1992) *The Inexact and Separate Science of Economics*, Cambridge, Cambridge University Press.

Hausman, D.M. and P. Mongin (1998) “Economists’ Responses to Anomalies: Full-Cost Pricing versus Preference Reversals”, in J. Davis (ed.) *New Economics and Its History*, HOPE Supplement Vol. 29, Durham, Duke University Press.

Hempel, C.G. (1965) *Aspects of Scientific Explanation*, New York, Free Press.

Hempel, C.G. (1988) “Provisos: A Problem Concerning the Inferential Function of Scientific Theories”, in A. Grünbaum and W.C. Salmon (eds.) *The Limitations of Deductivism*, Berkeley and Los Angeles, University of California Press.

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

Herstein, I. and J. Milnor (1953) "An Axiomatic Approach to Measurable Utility", *Econometrica*, 47, pp. 291-297.

Hesse, M.B. (1963) *Models and Analogies in Science*, London, Sheed and Ward.

Hey, J.D. (1991) *Experiments in Economics*, Oxford, Blackwell.

Hicks, J. and R.G.D. Allen (1934) “A Reconsideration of the Theory of Value”, *Economica*, 1, pp. 52-76, 196-219.

Holt, C.A. (1986) "Preference Reversals and the Independence Axiom", *American Economic Review*, 76, pp. 508-515.

Holton, G. (1978) *The Scientific Imagination*, Cambridge, Cambridge University Press.

Hon, G. (1989) “Towards a Typology of Experimental Error: An Epistemological View”, *Studies in History and Philosophy of Science*, 20, pp. 469-504.

Howson, C. (ed. 1976) *Method and Appraisal in the Physical Sciences*, Cambridge, Cambridge University Press.

Hughes, R.I.G. (1997) “Models and Representation”, *Philosophy of Science*, 64 (Proceedings), pp. S325-S336.

Hughes, R.I.G. (1999) “The Ising Model, Computer Simulation, and Universal Physics”, in M.S. Morgan and M.C. Morrison (eds.) *Models as Mediators*, Cambridge, Cambridge University Press.

Hume, D. (1740) *A Treatise of Human Nature*, Oxford, Clarendon Press, 1978.

Humphreys, P. (1989) *The Chances of Explanation*, Princeton, Princeton University Press.

Hutchison, T.W. (1938) *The Significance and Basic Postulates of Economic Theory*, New York, Kelley.

Jorland, G. (1987) “The Saint Petersburg Paradox 1713-1937”, in L. Krüger, L.J. Daston and M. Heidelberger (eds.) *The Probabilistic Revolution*, Vol. 1, Cambridge, Mass., MIT Press.

Kagel, J.H. and A.E. Roth (eds. 1995) *The Handbook of Experimental Economics*, Princeton, Princeton University Press.

Kagel, J.H. and R. Battalio (1980) “Token Economy and Animal Models for the Experimental Analysis of Economic Behavior”, in J. Kmenta and J. Ramsey (eds.) *Evaluation of Econometric Models*, New York, New York University Press.

Kagel, J.H. and D. Levin (1986) “The Winner’s Curse Phenomenon and Public Information in Common Value Auctions”, *American Economic Review*, 76, pp. 894-920.

Kagel, J.H., R.C. Battalio, and L. Green (1995) *Economic Choice Theory*, Cambridge, Cambridge University Press.

Kahneman, D. and A. Tversky (1979) “Prospect Theory: An Analysis of Decision under Risk”, *Econometrica*, 47, pp. 263-291.

Karni, E. and Z. Safra (1987) '"Preference Reversal' and the Observability of Preferences by Experimental Methods", *Econometrica*, 55, pp. 675-685.

Keller, L.R., U. Segal, and T. Wang (1993) “The Becker-DeGroot-Marschak Mechanism and Generalized Utility Theories: Theoretical Predictions and Empirical Observations”, *Theory and Decision*, 34, pp. 83-97.

Kincaid, H. (1990) “Defending Laws in the Social Sciences”, *Philosophy of the Social Sciences*, 20, pp. 56-83.

Kincaid, H. (1996) *Philosophical Foundations of the Social Sciences*, Cambridge, Cambridge University Press.

Kitcher, P. (1981) “Explanatory Unification”, *Philosophy of Science*, 48, pp. 507-531.

Klappholz, K. and J. Agassi (1959) “Methodological Prescriptions in Economics”, *Economica*, 26, pp. 60-74.

Knez, M. and V.L. Smith (1987) “Hypothetical Valuations and Preference Reversals in the Context of Asset Trading”, in A.E. Roth (ed.) *Laboratory Experimentation in Economics: Six Points of View*, Cambridge, Cambridge University Press.

Knight, F.H. (1922) *Risk, Uncertainty, and Profit*, Boston, Houghton Mifflin.

Knorr-Cetina, K. (1992) “The Couch, The Cathedral, And the Laboratory: On the Relationship between Experiment and Laboratory in Science”, in A. Pickering (ed.) *Science as Practice and Culture*, Chicago, University of Chicago Press.

Koestler, A. (1959) *The Sleepwalkers*, London, Hutchinson.

Kosso, P. (1989) “Science and Objectivity”, *Journal of Philosophy*, 86, pp. 245-257.

Krajewski, W. (1977) *Correspondence Principle and Growth of Science*, Dordrecht, Reidel.

Kreps, D.N. (1990) *A Course in Microeconomic Theory*, Princeton, Princeton University Press.

Kuhn, T.S. (1962/1970) *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press, 2nd ed.

Kuhn, T.S. (1974) "Second Thoughts on Paradigms" in F. Suppe (ed.) *The Structure of Scientific Theories*, Urbana, University of Illinois Press; reprinted in *The Essential Tension*, Chicago, University of Chicago Press, 1977.

LaFollette, H. and N. Shanks (1996) *Brute Science: Dilemmas of Animal Experimentation*, London, Routledge.

Lakatos, I. (1962) "Infinite Regress and the Foundations of Mathematics", reprinted in *Mathematics, Science and Epistemology: Philosophical Papers, Volume 2*, ed. by J. Worrall and G. Curry, Cambridge, Cambridge University Press, 1978.

Lakatos, I. (1963-64/1976) *Proofs and Refutations. The Logic of Mathematical Discovery*, edited by J. Worrall and E. Zahar, Cambridge, Cambridge University Press.

Lakatos, I. (1967) "A Renaissance of Empiricism in the Recent Philosophy of Mathematics", reprinted in *Mathematics, Science and Epistemology: Philosophical Papers, Volume 2*, ed. by J. Worrall and G. Curry, Cambridge, Cambridge University Press, 1978.

Lakatos, I. (1970) "Falsificationism and the Methodology of Scientific Research Programmes", in *The Methodology of Scientific Research Programmes: Philosophical Papers, Volume 1*, ed. by J. Worrall and G. Curry, Cambridge, Cambridge University Press, 1978.

Lakatos, I. (1971) "History of Science and Its Rational Reconstructions", in *The Methodology of Scientific Research Programmes: Philosophical Papers*, Vol. 1, ed. by J. Worrall and G. Currie, Cambridge, Cambridge University Press, 1978.

Larvor, B. (1998) *Lakatos: An Introduction*, London, Routledge.

Latour, B. (1986) *Science in Action*, Cambridge Mass., Harvard University Press.

Latour, B. and S. Woolgar (1979/1986) *Laboratory Life. The Construction of Scientific Facts*, Princeton, Princeton University Press, 2nd ed.

Latsis, S. (ed. 1976) *Method and Appraisal in Economics*, Cambridge, Cambridge University Press.

Laymon, R. (1985) "Idealization and the Testing of Theories by Experimentation", in P. Achistein and O. Hannaway (eds.)

Observation, Experiment, and Hypothesis in Modern Physical Science, Cambridge Mass., MIT Press.

Le Grand, H.E. (ed. 1990) *Experimental Inquiries*, Dordrecht, Kluwer.

Leonard, R. (1994) "Laboratory Strife: Higgling as Experimental Science in Economics and Social Psychology", in N. De Marchi and M.S. Morgan (eds.) *Higgling*, HOPE supplement, Vol. 26, Durham, Duke University Press.

Lichtenstein, S. and P. Slovic (1971) "Reversals of Preference Between Bids and Choices in Gambling Decisions", *Journal of Experimental Psychology*, 89, pp. 46-55.

Lichtenstein, S. and P. Slovic (1973) "Response-Induced Reversals of Preference in Gambling: An Extended Replication in Las Vegas", *Journal of Experimental Psychology*, 101, pp. 16-20.

Longworth, F. (unpublished) "Computer Modeling in the Earth Sciences", Department of Geology, Imperial College of Science, Technology and Medicine, London.

Loomes, G. (1989) "Experimental Economics", in J.D. Hey (ed.) *Current Issues in Microeconomics*, New York, St. Martin's Press.

Loomes, G. and R. Sugden (1982) "Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty", *Economic Journal*, 92, pp. 805-824.

Loomes, G. and R. Sugden (1983) "A Rationale for Preference Reversal", *American Economic Review*, 73, pp. 428-432.

Loomes, G. and R. Sugden (1984) "The Importance of What Might Have Been", in Hagen and Wenstøp, F. (eds. 1984) *Progress in Utility and Risk Theory*, Dordrecht, Reidel.

Loomes, G., C. Starmer and R. Sugden (1989) "Preference Reversal: Information-Processing Effect or Rational Non-Transitive Choice?", *Economic Journal*, 99, pp. 140-151.

Loomes, G., C. Starmer and R. Sugden (1991) "Observing Violations of Transitivity by Experimental Methods", *Econometrica*, 59, pp. 425-439.

Luce, R.D. and H. Raiffa (1957) *Games and Decisions*, New York, Wiley.

Lynch, M. (1985) *Art and Artifact in Laboratory Science*, London, Routledge.

Lynch, M. (1990) "The Externalised Retina", in M. Lynch and S. Woolgar (eds.) *Representation in Scientific Practice*, Cambridge Mass., MIT Press.

MacCrimmon, K. R. (1968) "Descriptive and Normative Implications of Decision Theory Postulate" in K. Borch and J. Mossin (eds. 1968) *Risk and Uncertainty*, London, St. Martin Press.

MacCrimmon, K.R. and S. Larsson (1979) "Utility Theory: Axioms vs. Paradoxes", in M. Allais and O. Hagen (eds.) *Expected Utility Hypothesis and the Allais Paradox*, Dordrecht, Reidel.

MacDonald, D.N., W.L. Huth, and P.M. Taube (1992) "Generalised Expected Utility Analysis and Preference Reversals: Some Initial Results in the Loss Domain", *Journal of Economic Behavior and Organization*, 17, pp. 115-130.

Machina, M. J. (1982) "'Expected Utility' Analysis without the Independence Axiom", *Econometrica*, 50, pp. 277-323.

Machina, M.J. (1983) "Generalized Expected Utility Analysis and the Nature of Observed Violations of the Independence Axiom", in B.P. Stigum and F. Wenstøp (eds.) *Foundations of Utility and Risk Theory with Applications*, Dordrecht, Reidel..

Marschak, J. (1950) "Rational Behaviour, Uncertain Prospects, and Measurable Utility", *Econometrica*, 18, pp. 111-141.

May, K.O. (1954) "Intransitivity, Utility and the Aggregation of Preference Patterns", *Econometrica*, 22, pp. 1-13.

Mayo, D. (1996) *Error and the Growth of Experimental Knowledge*, Chicago, University of Chicago Press.

McAfee, R.P. and J. McMillan (1987) "Auctions and Bidding", *Journal of Economic Literature*, 25, pp. 699-738.

McCloskey, D. (1985) *The Rhetoric of Economics*, Madison, University of Wisconsin Press.

McMullin, E. (1977) "Material Causality", in R. Butts and J. Hintikka (eds.) *Historical and Philosophical Dimensions of Logic, Methodology and Philosophy of Science*, Dordrecht, Reidel.

McMullin, E. (1985) "Galilean Idealization", *Studies in History and Philosophy of Science*, 16, pp. 247-273.

Mead, W.J., A. Moseidjord and P.E. Sorensen (1983) "The Rate of Return Earned by Leases Under Cash Bonus Bidding in the OCS Oil and Gas Leases", *Energy Journal*, 4, pp. 37-52.

Milgrom, P. (1989) “Auctions and Bidding: A Primer”, *Journal of Economic Perspectives*, 3, pp. 3-22.

Milgrom, P.R. and R.J. Weber (1982) “A Theory of Auctions and Competitive Bidding”, *Econometrica*, 50, pp. 1089-1122.

Mill, J.S. (1836) “On the Definition of Political Economy ad the Method of Investigation Proper to It”, in *Collected Works of John Stuart Mill,* Vol. 4, Toronto, University of Toronto Press, 1967.

Mill, J.S. (1843) *A System of Logic*, London, Longman Green & Co., 1949.

Mirowski, P. (1989) *More Heat than Light*, Cambridge, Cambridge University Press.

Mirowski, P. (forthcoming) *Machine Dreams*, Cambridge Mass., Harvard University Press.

Mongin, P. (1988a) "Problèmes de Duhem en théorie de l'utilité espérée", *Fundamenta Scientiae*, 9, pp. 299-327.

Mongin, P. (1988b) “Le réalisme des hypothèse et la *Partial Interpretation View*”, *Philosophy of the Social Sciences,* 18, pp. 281-325.

Mongin, P. (1995) “L’utilitarisme originel et le développement de la théorie économique”, *La Pensée Politique*, 3, pp. 341-361.

Mongin, P. and C. d’Aspremont (1996) “Utility Theory and Ethics”, in S. Barberà, P.J. Hammond and C. Seidl (eds.) *Handbook of Utility Theory,* Kluwer, Amsterdam

Morgan, M.S. (1990) *The History of Econometric Ideas,* Cambridge, Cambridge University Press.

Morgan, M.S. and M. Boumans (1998) “The Secrets Hidden by Two-Dimansionality: Modelling the Economy as a Hydraulic System”, *Research Memoranda in History and Methodology of Economics,* 98-2, University of Amsterdam.

Morgan, M.S. and M.C. Morrison (1999) “Models as Mediating Instruments”, in M.S. Morgan and M.C. Morrison (eds.) *Models as Mediators*, Cambridge, Cambridge University Press.

Morgenstern, O. (1972) "Descriptive, Predictive and Normative Theory", *Kyklos*, 4, pp. 699-714.

Morgenstern, O. (1979) "Some Reflections on Utility", in M. Allais and O. Hagen (eds.) *Expected Utility Hypothesis and the Allais Paradox*, Dordrecht, Reidel.

Morrison, M.C. (1990) “Theory, Intervention and Realism”, *Synthèse*, 82, pp. 1-22.

Morrison, M.C. (1998a) “Experiment”, in E. Craig (ed.) *The Routledge Encyclopedia of Philosophy*, London, Routledge.

Morrison, M.C. (1998b) “Mediating Models: Between Physics and the Physical World”, *Philosophia Naturalis*, 35.

Motterlini, M. (1999) *Imre Lakatos. Matematica, scienza e storia*, Milano, Il Saggiatore.

Mulkay, M. and G.N. Gilbert (1986) “Replication and Mere Replication”, *Philosophy of the Social Sciences*, 16, pp. 21-37.

Musgrave, A. (1981) “‘Unreal Assumptions’ in Economic Theory: The F-Twist Untwisted”, *Kyklos*, 34, pp. 377-387.

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

Nelson, A. (1986) “New Individualist Foundations for Economics”, *Nous*, 20, pp. 469-490.

Neurath, O. (1934) “Radikaler Physikalismus und ‘wirkliche Welt’”, *Erkenntnis*, 4, pp. 346-362; Engl. tr.. “Radical Physicalism and the ‘Real World’”, in *Philosophical Papers 1913-1946*, ed. and transl. by R.S. Cohen and M. Neurath, Dordrecht, Reidel, 1983.

Newton-Smith, W. (1981) *The Rationality of Science*, London, Routledge.

Nowakowa, I. (1994) *Idealization V: The Dynamics of Idealization*, Posznan Studies in the Philosophy of Science, vol. 34, Amsterdam, Rodopi.

Nye, M.J. (1972) *Molecular Reality*, London, Macdonald.

Pareto, V. (1906) *Manuale di economia politica*, Milano, Società Editrice Libraria.

Pareto, V. (1907) “L’économie et la sociologie au point de vue scientifique”, *Rivista di scienza*, 1907, 293-312; Italian transl. in *Oeuvres completes*, vol. 22: *Écrites sociologiques mineurs*, ed. by G. Busino, Genève, Librairie Droz, 1980.

Paul, J. (1971) *A History of Poliomelytis*, New Haven, Yale University Press.

Pickering, A. (1984) *Constructing Quarks*, Chicago, University of Chicago Press.

Pickering, A. (1992) "From Science as Knowledge to Science as Practice", in Pickering (ed.) *Science as Practice and Culture*, Chicago, University of Chicago Press.

Pickering, A. (1995a) "Beyond Constraint: The Temporality of Practice and the Historicity of Knowledge", in Buchwald (ed.) *Scientific Practice*, Chicago, University of Chicago Press.

Pickering, A. (1995b) *The Mangle of Practice: Time, Agency, and Science*, Chicago, University of Chicago Press.

Pickering, A. (ed. 1992) *Science as Practice and Culture*, Chicago, University of Chicago Press.

Plott, C.R. (1987) "Dimensions of Parallelism: Some Policy Applications of Experimental Methods", in A.E. Roth (ed.) *Laboratory Experimentation in Economics: Six Points of View*, Cambridge, Cambridge University Press.

Plott, C.R. (1991) "Will Economics Become an Experimental Science?" *Southern Economic Journal*, 57, pp. 901-919.

Plott, C.R. (1995) "Rational Individual Behaviour in Markets and Social Choice Processes: The Discovered Preference Hypothesis", in K.J. Arrow, E. Colombatto, M. Perlman and C. Schmidt (eds.) *The Rational Foundations of Economic Behaviour*, London, Macmillan.

Plott, C.R. (1997) "Laboratory Experimental Testbeds: Application to the PCS Auction", *Journal of Economics and Management Strategy*, 6, pp. 605-638.

Poincaré, H. (1905) *La valeur de la science*, Paris, Flammarion; Engl. transl. *The Value of Science*, Dover, Dover Publications.

Polanyi , M. (1967) *The Tacit Dimension*, New York, Anchor.

Pommerehne, W.W., F. Schneider and P. Zweifel (1982) "Economic Theory of Choice and the Preference Reversal Phenomenon: A Reexamination", *American Economic Review*, 72, pp. 569-574.

Popper, K.R. (1934/1959) *Logik der Forschung*, Vienna, Springer; Engl. transl. *Logic of Scientific Discovery*, London, Hutchinson.

Popper, K.R. (1957) “The Aim of Science”, *Ratio*, 1, pp. 24-35; reprinted in *Objective Knowledge*, Oxford, Clarendon Press, 1972.

Popper, K.R. (1963) *Conjectures and Refutations*, London, Routledge.

Popper, K.R. (1967) "The Rationality Principle", in *Popper Selections*, ed. by D. Miller, Princeton, Princeton University Press, 1985.

Post, H.R. (1971) “Correspondence, Invariance and Heuristics”, *Studies in History and Philosophy of Science*, 2, pp. 213-255.

Pratt, J.W. (1964) "Risk Aversion in the Small and in the Large", *Econometrica*, 32, pp. 122-136.

Putnam, H. (1974) “On the ‘Corroboration’ of Theories”, in P. Schilpp (ed.) *The Philosophy of Karl Popper*, La Salle, Ill., Open Court.

Quiggin, J. (1982) “A Theory of Anticipated Utility”, *Journal of Economic Behavior and Organization*, 3, pp. 323-343.

Quine, W.O. (1953) “Two Dogmas of Empiricism”, in *From A Logical Point of View*, Cambridge Mass., Harvard University Press.

Radder, H. (1992) “Experimental Reproducibility and the Experimenter’s Regress”, in *PSA 1992*, East Lansing: Philosophy of Science Association.

Radder, H. (1996) *In and About the World*, Albany, SUNY Press.

Ramsey, F. (1926) "Truth and Probability", in his *The Foundations of Mathematics and Other Logical Essays*, ed. by R.B. Braithwaite, London, Routledge, 1931.

Redhead, M. (1980) “Models in Physics”, *British Journal for the Philosophy of Science*, 31, pp. 145-163.

Reilly, R.J. (1982) “Preference Reversal: Further Evidence and Some Suggested Modifications in Experimental Design”, *American Economic Review*, 72, pp. 576-584.

Reiss, J. (unpublished) “The Path from Price Data to a Measure of Inflation”, CPNSS, London School of Economics.

Robbins, L. (1932) *An Essay on the Nature and Significance of Economic Science*, London, Macmillan.

Rorty, R. (1979) *Philosophy and the Mirror of Nature*, Princeton, Princeton University Press.

Rosenberg, A. (1992) *Economics: Mathematical Politics or Science of Diminishing Returns?* Chicago, University of Chicago Press.

Roth, A.E. (1986) "Laboratory Experimentation in Economics", *Economics and Philosophy*, 2, 245-273.

Roth, A.E. (1988) "Laboratory Experimentation in Economics: A Methodological Overview", *Economic Journal*, 98, 974-1031.

Roth, A.E. (1995) "Introduction to Experimental Economics", in J.H. Kagel and A.E. Roth (eds.) *The Handbook of Experimental Economics*, Princeton, Princeton University Press.

Roth, A.E. and M.W.K. Malouf (1979) "Game-Theoretic Models and the Role of Information in Bargaining", *Psychological Review*, 86, pp. 574-594.

Russell, B. (1912) "On Induction", in *The Problems of Philosophy*, Oxford, Oxford University Press, 1973.

Safra, Z., U. Segal and A. Spivak (1990a) "Preference Reversals and Non-Expected Utility", *American Economic Review*, 80, pp. 922-930.

Safra, Z., U. Segal, and A. Spivak (1990b) "The Becker-DeGroot-Marschak Mechanism and Non-Expected Utility: A Testable Approach", *Journal of Risk and Uncertainty*, 3, 177-190.

Sainsbury, R.M. (1988) *Paradoxes*, Cambridge, Cambridge University Press.

Salanti, A. (1994) "On the Lakatosian Apple of Discord in the History and Methodology of Economics", *Finnish Economic Papers*, 7, pp. 30-41.

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

Samuelson, P.A. (1950) "Probability and the Attempts to Measure Utility", *Economic Review*, July, pp. 169-170.

Samuelson, P.A. (1952) "Utilité, préference et probabilité", in C.N.R.S. (1953) *Économétrie* (Paris 12-17 mai 1952), Colloques Internationaux du Centre National de la Recherche Scientifique; reprinted as "Utility, Preference, and Probability", in *The Collected Scientific Papers of Paul A. Samuelson, Vol. 1*, ed. by J. Stiglitz, Cambridge (Mass.), MIT Press.

Savage, L.J. (1951) "The Theory of Statistical Decision", *Journal of the American Statistical Association*, 46, pp. 55-67.

Savage, L.J. (1952) "Une axiomatisation de comportement raisonnable face à l'incertitude", in C.N.R.S. (1953) *Économétrie* (Paris 12-17 mai

1952), Colloques Internationaux du Centre National de la Recherche Scientifique.

Savage, L.J. (1954) *The Foundations of Statistics*, New York, Dover Publications, 2nd edition 1972.

Schelling, T.C. (1978) *Micromotives and Macrobehavior*, New York, Norton.

Segal, U. (1988) "Does the Preference Reversals Phenomenon Necessarily Contradict the Independence Axiom?", *American Economic Review*, 28, pp. 175-202.

Sen, A. (1985) "Rationality and Uncertainty", *Theory and Decision*, 18, pp. 109-127.

Shapin, S. (1988) "The House of Experiment in Seventeenth-Century England", *Isis*, 79, pp. 373-404.

Shapin, S. and S. Schaffer (1985) *Leviathan and the Air-Pump*, Princeton, Princeton University Press.

Simon, H.A. (1969) *The Sciences of the Artificial*, Boston, MIT Press.

Slovic, P. (1995) "The Construction of Preferences", *American Psychologist*, 50, pp. 364-371.

Slovic, P. and A. Tversky (1974) "Who Accepts Savage's Axiom?" *Behavioural Science*, 19, pp. 368-373.

Slovic, P. and S. Lichtenstein (1968) "Relative Importance of Probabilities and Payoffs in Risk-Taking", *Journal of Experimental Psychology,* Monograph Supplement, Part 2, pp. 1-18.

Smith, V.L. (1962) "An Experimental Study of Competitive Market Behavior", *Journal of Political Economy,* 70, pp. 111-137; reprinted in Smith (1991).

Smith, V.L. (1976) "Experimental Economics: Induced Value Theory", *American Economic Review*, 66, pp. 274-277; reprinted in Smith (1991).

Smith, V.L. (1981) "Experimental Economics at Purdue", in G. Horwich and J.P. Quirk (eds.) *Essays in Contemporary Fields of Economics*, West Lafayette, Purdue University Press; reprinted in Smith (1991).

Smith, V.L. (1982) "Microeconomic Systems as an Experimental Science", *American Economic Review*, 72, pp. 923-955; reprinted in Smith (1991).

Smith, V.L. (1987) “Experimental Methods in Economics”, in J. Eatwell, M. Milgate, and P. Newman (eds.) *The New Palgrave*, Vol. 2, London, Macmillan.

Smith, V.L. (1989) “Theory, Experiment and Economics”, *Journal of Economic Perspectives,* 3, pp. 151-169; reprinted in Smith (1991).

Smith, V.L. (1991) *Papers in Experimental Economics*, Cambridge, Cambridge University Press.

Smith, V.L. (1992) “Game Theory and Experimental Economics: Beginnings And Early Influences”, in E.R. Weintraub (ed.) *Towards A History of Game Theory*, *History of Political Economy*, supplement to Vol. 24, Durham, Duke University Press.

Smith, V.L. (1994) “Economics in the Laboratory”, *Journal of Economic Perspectives*, 8, pp. 113-131.

Smith, V.L., K.A. McCabe and S.J. Rassenti (1991) “Lakatos and Experimental Economics”, in N. de Marchi and M. Blaug (eds.) *Appraising Economic Theories*, Aldershot, Elgar.

Starmer, C. (1999) “Experiments in Economics... (Should We Trust the Dismal Scientists in White Coats?)”, *Journal of Economic Methodology,* 6, pp. 1-30.

Starmer, C. (forthcoming) “Developments in Non-Expected Utility Theory: The Hunt for A Descriptive Theory of Choice Under Risk”, *Journal of Economic Literature*.

Starmer, C. and R. Sugden (1991) “Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation”, *American Economic Review*, 81, pp. 971-978.

Stegmüller, W. (1979) *The Structuralist View of Science,* New York, Springer Verlag.

Steinle, F. (1997) “Entering New Fields: Exploratory Uses of Experimentation”, *Philosophy of Science,* 64, pp. S65-S74.

Stuwer, R.H. (1985) “Artificial Disintegration and the Vienna-Cambridge Constroversy”, in P. Achistein and O. Hannaway (eds.) *Observation, Experiment and Hypothesis in Modern Physical Science*, Cambridge Mass., MIT Press.

Sugden, R. (1985) "Regret, Recrimination and Rationality", *Theory and Decision*, 19, pp. 77-99.

Sugden, R. (1991) "Rational Choice: A Survey of Contributions from Economics and Philosophy", *Economic Journal*, 101 (407), pp. 751-785.

Sugden, R. (1993) "An Axiomatic Foundation for Regret Theory", *Journal of Economic Theory* 60: 159-80.

Sugden, R. (forthcoming) "Credible Worlds: The Status of Theoretical Models in Economics", *Journal of Economic Methodology*.

Suppe, F. (1989) *The Semantic Conception of Theories and Scientific Realism*, Urbana, University of Illinois Press.

Suppes, P. (1962) "Models of Data", in E. Nagel, P. Suppes and A. Tarski (eds.) *Methodology and Philosophy of Science*, Stanford, Stanford University Press; reprinted in Suppes (1969).

Suppes, P. (1969) *Studies in the Methodology and Foundations of Science*, Dordrecht, Reidel.

Suppes, P. (1981) "The Plurality of Science", in P. Asquith and I. Hacking (eds.) *PSA 1978*, Vol. 2, East Lansing MI, Philosophy of Science Association.

Suppes, P. (1984) *Probabilistic Metaphysics*, London, Blackwell.

Sutton, J. (forthcoming) *Marshall's Tendencies. What Can Economists Know?*

Tammi, T. (1997) *Essays on the Rationality of Experimentation in Economics*, University of Joensuu Publications in Social Sciences, 27.

Thaler, R.H. (1988) "The Winner's Curse", *Journal of Economic Perspectives*, 2, pp. 191-202.

Thaler, R.H. and A. Tversky (1990) "Preference Reversals", *Journal of Economic Perspectives*, 4, pp. 201-211.

Tversky, A. (1969) "Intransitivity of Preferences", *Psychological Review*, 76, pp. 31-48.

Tversky, A. and R.H. Thaler (1990) "Preference Reversals", *Journal of Economic Perspectives*, 4, pp. 201-211.

Tversky, A., P. Slovic, and D. Kahneman (1990) "The Causes of Preference Reversals", *American Economic Review*, 80, pp. 204-217.

van Fraassen, B. (1980) *The Scientific Image*, Oxford, Oxford University Press.

Vickrey, W. (1961) "Counterspeculation, Auctions, and Competitive Sealed Tenders", *Journal of Finance*, 16, pp. 8-37.

Von Neumann, J. and O. Morgenstern (1944) *The Theory of Games and Economic Behavior*, Princeton, Princeton University Press.

Ward, B. (1972) *What's Wrong with Economics?*, London, Macmillan.

Watkins, J. (1984) *Science and Scepticism*, Princeton, Princeton University Press.

Weldon, J.C. (1950) "A Note on Measures of Utility", *Canadian Journal of Economics and Political Science*, 16, 227-233.

Weldon, J.C. (1953) "Discussion", *Econometrica*, 21, p. 350.

Wilde, L.L. (1981) "On the Use of Laboratory Experiments in Economics", in J.C. Pitt (ed.) *Philosophy in Economics*, Dordrecht, Reidel.

Will, C. (1988) *Was Einstein Right?*, Oxford, Oxford University Press.

Wilson, R. (1977) "A Bidding Model of Perfect Competition", *Review of Economic Studies*, 44, pp. 511-518.

Woodward, J. (1989) "Data and Phenomena", *Synthèse*, 79, pp. 393-472.

Woodward, J. (unpublished) "Explanation and Invariance in the Special Sciences", Department of Philosophy, California Institute of Technology.

Worrall, J. (1978) "The Ways in which the Methodology of Scientific Research Programmes Improves on Popper's Methodology", in G. Andersson and A. Radnitzky (eds.) *Progress and Rationality of Science*, Boston Studies in the Philosophy of Science, Dordrecht, Reidel.

Worrall, J. (1985) "Scientific Discovery and Theory-Confirmation", in J. Pitt (ed.) *Change and Progress in Modern Science*, Dordrecht, Reidel.

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

Worrall, J. (1991) "Feyerabend and the Facts", in G. Munévar (ed.) *Beyond Reason: Essays on the Philosophy of Paul K. Feyerabend*, Dordrecht, Kluwer.

Yaari, M.E. (1987) "The Dual Theory of Choice Under Risk", *Econometrica*, 55, pp. 95-115.

Zahar, E. (1976) "Why Did Einstein's Programme Supersede Lorentz's", in C. Howson (ed.) *Method and Appraisal in the Physical Sciences*, Cambridge, Cambridge University Press.

Zahar, E. (1983) “Logic of Discovery or Psychology of Invention?”, *British Journal for the Philosophy of Science*, 34, pp. 243-261.

Zahar, E. (1997) *Leçons d’épistemologie*, in *Cahiers du CREA*, ed. by A. Boyer, Paris, École Polytechnique.